MEASURING THE PHYSICAL STEP-SIZE OF BACTERIOPHAGE T7 DNA HELICASE WITH DUAL OPTICAL TWEEZERS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Benjamin Yuri Smith

January 2011

MEASURING THE PHYSICAL STEP-SIZE OF BACTERIOPHAGE T7 DNA HELICASE WITH DUAL OPTICAL TWEEZERS

Benjamin Yuri Smith, Ph. D.

Cornell University 2011

Helicases are ubiquitous molecular motors which unwind double stranded DNA into two single strands. T7gp4 is a model hexameric helicase from the replication system of bacteriophage T7. The step size of T7gp4 was probed using single molecule optical trapping experiments. To obtain sufficient resolution over experimental timescales a dual optical trap was constructed, decoupling the experiment from the surface. Numerical simulations of helicase in the experimental configuration were conducted to examine the required measurement bandwidth. A novel method of forming double tethers between the dual traps was implemented. A range of steps of the helicase were observed. In combination with structural data and biochemical measurements, this agrees with a sequential mechanism for T7gp4 translocation.

BIOGRAPHICAL SKETCH

Benjamin Yuri Smith was born in London, England in 1981. He attended Fairlawn, Forest Hill Boys, and Hillsyde sixth form. He then studied at Bristol University where he earned a Masters of Science in Physics in 2003. After being accepted to the graduate program at Cornell University, he moved to the United States and began working with Professor Carl Franck in the Physics department in 2004. In 2006 he spent a summer working with Professor Carl Batt, before joining the lab of Michelle Wang. He graduated with a Masters of Science degree in 2009, and a Ph.D. in 2010.

To my grandparents Anna and Rose, and my great-aunt Laura

ACKNOWLEDGMENTS

First I would like to thank my Ph.D. supervisor Professor Michelle Wang for all of her scientific advice, and the rigorous training which is a hallmark of her lab. Not only has my time in her lab trained me to be a good scientist, it has also taught me how to communicate ideas and results clearly and effectively.

I would also like to thank our collaborator and friend Professor Smita Patel at the Robert Wood Johnson Medical School, as well as members of her lab. She provided us with the proteins without which none of this work would be possible, and as one of the world's experts on T7 helicase she helped us immensely in interpreting and understanding our results.

I would also like to thank all the members of the Wang lab, who every day have provided scientific insight, advice, support and encouragement. In particular I would like to thank Daniel Johnson who developed many of the protocols used in this work, and trained me to perform helicase experiments. I would also like to thank Chris Deufel for guiding me through the instrumentation learning process, and Lucy Bai for introducing me to single-molecule simulations and theory. Additionally I would like to thank Michael Hall, Scott Forth and James Inman for discussions about instrumentation, and Jing Jin, Bert Fulbright and Maxim Sheinin for help with molecular biology. I would also like to thank Bo Sun for help and discussions regarding the helicase assay.

I would like to thank my special committee members, James Sethna and Chris Xu, and also Lois Pollack who stood in for Chris during my thesis defense.

Finally, I would like to thank my family for being supportive and loving and encouraging my pursuits and interest in science – and everything else – throughout my life. My parents, Emmy, David, Digby and Subrena; my grandparents Anna, Rose and Laura; My sister and friend Sasha; and the two women who make every day a pleasure: my wife Gwen, and my cat Beatrix.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1:
Introduction

Helicases

Helicases are ubiquitous molecular motors which unwind double stranded DNA (dsDNA) into two strands of single stranded DNA (ssDNA) (S.S. Patel & K.M. Picha 2000). They extract energy from hydrolysis of nucleotide triphosphates (NTPs) which fuel the translocation and unwinding activity (Lohman & Bjornson 1996). Helicases are vital for all forms of life. They perform fundamental roles in repair, replication, translation, transcription and recombination (Alberts 2007). Understanding the mechanisms that helicases use in order to accomplish their biological function provides us with a more complete knowledge of one of the most basic and important classes of proteins.

Helicases separate the two strands of dsDNA by breaking the hydrogen bonds which link the complementary bases (Alberts 2007). The separation of dsDNA is important for a variety of processes in the cell, and a diverse selection of helicases has been identified, each evolved to fill their specific biological niche.

For example, the helicase DnaB is responsible for opening the replication fork in *E. coli* (A. Johnson & O'Donnell 2005). This is necessary because the molecular machinery responsible for duplicating the genome, *E. coli* Pol III, cannot read the genetic code while it is in the form of dsDNA. The helicase is responsible for converting the more stable, less accessible dsDNA into ssDNA which can be read by the rest of the replication machinery. In this case the protein function of the helicase and polymerase proteins is so inextricably linked that they form an interconnected complex during cell replication. This combination of proteins is called the replisome and occurs in organisms ranging from viruses to humans, but with a variety of

different helicases, polymerases and primases taking on the specific roles (Pomerantz & Odonnell 2007).

Another example of helicase function is during DNA nucleotide-excision repair (Alberts 2007). This is the biological process which repairs DNA bases that have become damaged, for example by ultraviolet radiation (Howard-Flanders 1968). Various proteins are recruited to the damaged site, at least one of which will be a helicase, in order to separate the damaged bases from the undamaged strand. Subsequently, a polymerase will fill in the removed bases and a ligase will join the repaired section to the existing DNA (Alberts 2007). For example, the helicase UvrD is an integral part of the repair process in *E. coli* (Grossman et al. 1988).

Other examples of helicase function include chromatin remodeling in eukaroyotes by the Tip48 helicase (Puri et al. 2007); mitochondrial DNA maintenance by Twinkle (Tyynismaa et al. 2004) and RNA processing by the DEAD-box helicase elF4A (Rogers et al. 2002).

Not only are helicases an important family of proteins in their own right, but they also share certain broadly similar structural motifs with a larger group of proteins. One example is the RecA motif present in many helicases (Ye et al. 2004). The RecA fold is named after the repair protein from *E. coli* it was first identified in (Story et al. 1992). The RecA fold forms a site in which NTPs are hydrolyzed and converted into a conformational change of the protein. It is therefore a biological version of a combustion cylinder, converting a chemical energy source into mechanical motion. The RecA motif is also present in a diverse selection of other proteins (Ye et al. 2004). These proteins range from helicases, to the more general class of proteins which move

directionally along DNA called translocases, to the even more general class known as ATPases (Singleton et al. 2007). All proteins in which the RecA fold occurs are multimeric, because an important feature of the motif is that there must be at least two RecA folds present in order for hydrolysis to occur (Singleton & Wigley 2003). Some examples of proteins containing the RecA fold are shown in Table 1. By studying just one particular example of these proteins we can potentially understand a much wider class of enzymes, particularly with regards to hydrolysis and co-operativity.

Table 1) Examples of proteins containing a RecA fold

Protein	**Number of RecA domains**	**Function**
PcrA	2	DNA repair helicase
Clamp loader	5	Loading protein for DNA polymerase
T7 Helicase	6	Replicative helicase
F1 ATPase	6	ATP synthesis
NSF	10	Membrane fusion

Classifying Helicases

Helicases can be broadly categorized into two distinct groups: ring-shaped and non-ring-shaped, depending on how many of the individual subunits are required to form a fully functional protein (Lohman & Bjornson 1996). Non-ring-shaped helicases require only one or two subunits to function, and are also known as monomeric/dimeric helicases, whereas ring-shaped helicases are formed from a donut shaped ring of subunits. Typically these rings are hexameric, although heptameric oligomers also are observed (Crampton et al. 2006). The subunits in a helicase may be

identical such as in T7gp4 (S S Patel & M M Hingorani 1993), or non-identical as in MCM helicases (Costa & Onesti 2008).

These groupings can be further subdivided into superfamilies based on the underlying amino acid sequence. Because the superfamilies are divided based on sequence there are some functional similarities between helicases in different superfamilies, and functional differences between helicases within a superfamily. Superfamilies I and II contain the non-ring-shaped helicases, and superfamilies III-VI the ring-shaped helicases (Gorbalenya & Koonin 1993).

Another method of organizing helicases into groups is by characterizing their functionality. One proposed set of groupings is based on five important parameters which can be used to describe the helicase mechanism (Singleton et al. 2007).

First, each helicase has a certain polarity. It will translocate in a certain direction when bound to a single strand of DNA. This may either be 5' → 3', such as the superfamily 1B helicases, or 3' → 5' such as the superfamily 1A helicases (Saikrishnan et al. 2009).

Second, each helicase translocates and unwinds at a different rate. For example, under physiological conditions T7 helicase translocates at around 300nt/s on an ssDNA substrate at 23 Celsius (D. S. Johnson et al. 2007). This rate is sensitive to a variety of factors such as DNA sequence, nucleotide concentration and buffer conditions. The presence of other proteins can also affect the rate. For example, T7gp4 was found to unwind dsDNA ten times faster in the presence of T7 polymerase (Stano et al. 2005).

Third, each helicase exhibits a certain processivity. It will tend to fall off a substrate after a certain amount of activity. The processivity will be affected by other proteins and the experimental conditions, and can range from a few, to tens of thousands of basepairs (Lohman & Bjornson 1996).

Fourth, each helicase exhibits a certain degree of activeness. This corresponds to how much the double stranded DNA is encouraged to separate into ssDNA strands (M D Betterton & Jülicher 2005). It could be that the helicase is entirely passive and functions as a Brownian ratchet. In this case it will not move unless thermal fluctuations open the DNA fork by the requisite number of basepairs for a step to be taken. On the other hand, a completely active helicase applies a force to the junction large enough to break the hydrogen bonds of the DNA. In general the activeness of a helicase will fall somewhere in between these extremes. The activeness can be precisely described by specifying the shape and magnitude of the applied potential.

Fifth, each helicase has a certain step size. Helicases do not move in a continuous fashion, but in discrete steps, relating to the conformational change in the protein which generates the forward motion and the mechanism of binding to the DNA substrate. This physical step may be different from the biochemical, or rate-limiting step (Smita S. Patel & Donmez 2006).

This classification system is powerful because it reveals a lot of information about the particular protein. If these parameters can be determined for a specific helicase, particularly in combination with knowledge of the structure, we have a good understanding of that enzyme.

T7 DNA Helicase

T7 DNA helicase is a model hexameric helicase in superfamily IV (Gorbalenya & Koonin 1993). It is expressed during the replication cycle of the T7 bacteriophage. It has been extensively studied for over thirty years and is particularly attractive because of the relatively simple replication system of the bacteriophage T7 (Scherzinger & Seiffert 1975). It is one of only four proteins required for T7 phage to successfully replicate its genome. The others are T7 polymerase, single-stranded binding protein, and thioredoxin (Lee et al. 1998). The helicase is required to separate the double-stranded genomic DNA into two single strands which can be accessed by the polymerase.

It is thought to accomplish strand separation by binding to an exposed region of ssDNA so that the strand passes through the central channel formed by the ring (Ahnert & S S Patel 1997). This central channel is too small to fit two strands of DNA, so as the helicase translocates along the single strand the second strand is displaced and the unwinding junction is advanced (Toth et al. 2003). T7gp4 has a 5'-> 3' polarity (E H Egelman et al. 1995). This is shown schematically in Figure 1.

The gene encoding T7 DNA helicase is known as gene product 4 (gp4), and it encodes two versions of the enzyme with different start codons. The longer protein (63kD) is a bifunctional helicase/primase and is known as gp4A, whereas the slightly shorter (56kD) protein functions only as a helicase and is called gp4B (Bernstein & C C Richardson 1988).

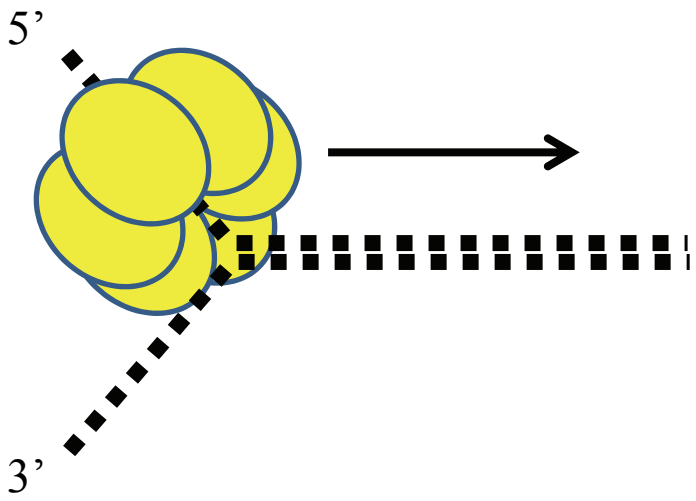

Figure 1) Model of T7 Helicase/Primase unwinding dsDNA into two strands of ssDNA.

The bifunctional gp4A protein contains both helicase and primase domains, flexibly combined via a covalent linker, forming two stacked rings (Mendelman & C C Richardson 1991). When translocating along ssDNA the helicase domain precedes the primase domain, which primes the lagging strand ssDNA as the helicase separates it from the complimentary strand (Benkovic et al. 2001).

In vivo the helicase/primase forms complexes with two units of T7 DNA polymerase forming the viral replisome (Benkovic et al. 2001). The helicase acts to separate the double stranded DNA ahead of the complex, advancing the replication fork. Each of the polymerase molecules acts on one of the two strands. In this manner the entire viral genome is replicated. T7gp4 is thus vital for the propagation of the virus.

In order to translocate and unwind T7gp4 requires free NTPs as an energy source, as well as a metal ion cofactor (Kristen Moore Picha & Smita S. Patel 1998). Helicase can form rings around ssDNA without the presence of a cofactor, but will not exhibit translocation or unwinding activity. dTTP has traditionally been used to conduct experiments since the processivity is high with this nucleotide and the binding is strongest (Matson & C C Richardson 1983). NTP's are hydrolyzed in each of six RecA folds which occur between each pair of subunits in the ring.

The ring which forms the complete T7gp4 seems to accommodate either six or seven subunits (Toth et al. 2003). The heptameric crystal structure is shown in Figure 2. However, it has traditionally been referred to as a hexamer because that form has been the most prevalent when observed both with native gel separations (S S Patel & M M Hingorani 1993) and with electron microscopy (E H Egelman et al. 1995). There is increasing evidence that T7gp4 and other so-called hexameric helicases may actually exist in both states (Miyata et al. 2000). It has been proposed that the heptameric form of T7gp4 may be able to translocate along dsDNA since the channel is wide enough to accommodate both strands and there is some experimental evidence for this process (Kaplan & O'Donnell 2002).

Each subunit contains loops which bind to the DNA substrate, facing into the central channel (Sawaya et al. 1999). A crystal structure of a shorter version of the protein complexed with the non-hydrolyzable NTP analog ADPNP indicated that the NTP hydrolysis state affects the relative rotation between subunits (Singleton et al. 2000). If this rotation is responsible for translocation along the substrate it indicates a step-size of 1-2bp based on the dimensions of the crystal structure. Additionally, biochemical assays suggest a physical step-size of 1-2bp on GC rich constructs, and 4bp on AT

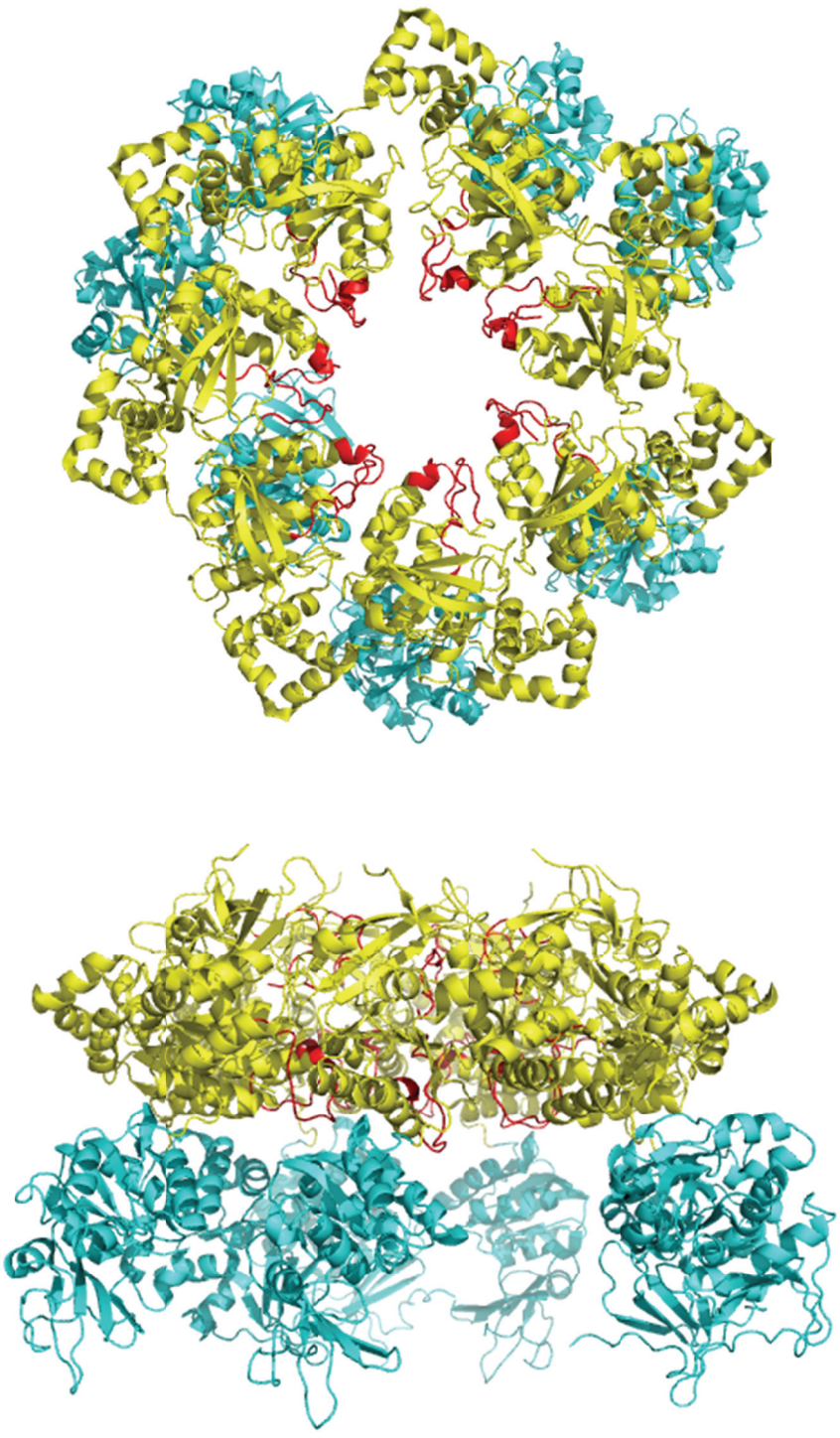

Figure 2) Crystal structure of T7gp4 helicase (Toth et al. 2003). The helicase domains are colored yellow, whereas the primase domains are shown in blue. Red indicates the DNA binding loops.

rich constructs (Donmez & Smita S Patel 2008). Single molecule studies indicate a likely step-size of ~2-4bp on a mixed construct (D. S. Johnson et al. 2007). Biochemical studies also indicate a hydrolysis rate of 1 dTTP for every 2-3bp of DNA unwound (Kim et al. 2002). There is also evidence to suggest that a rate limiting step occurs every 6-9bp (Jeong et al. 2004).

To summarize the structural data, T7gp4 consists of either six or seven subunits organized into a toroid. ssDNA fits through the central channel and can be bound to any of six binding loops which protrude into the center of the ring. The orientation of these binding loops is dependent on the hydrolysis state of the NTP in the hydrolysis pocket between subunits. This immediately suggests a variety of possible models for translocation on DNA. These will be examined in the following section.

Models for hexameric helicase translocation

A full understanding of T7gp4, or any helicase, would explain every step in the transformation of energy in chemical form, stored as NTPs, into the mechanical form used to translocate on and separate strands of DNA. This is referred to as mechano-chemical coupling. It would also account for all of the interactions with nucleic acids and related proteins.

One of the most important aspects of this process for a hexameric helicase is to understand how the subunits cooperate and interact with one another. For example, it may be that each subunit acts independently from the rest. This stochastic model would imply that each subunit would hydrolyze NTPs at its own rate, independent of the state of any other subunit. This is unlikely to be the case with T7gp4 as even a

small amount of non-functional monomer added in poisoning experiments significantly reduces the activity of the protein (Liao et al. 2005).

An opposing mechanism is complete cooperativity between subunits. It could be that all the subunits undergo a power stroke simultaneously, every arm swiveling in synchronicity. Each subunit would be unable to hydrolyze NTPs until all the other subunits were ready. This model conflicts with the structural data showing adjacent subunits in different states for T7gp4 (Singleton et al. 2000), but has been proposed as a mechanism for SV40 (Gai et al. 2004).

Another alternative is a grouped sequential model. This is similar to F1-ATPase in which the subunits are grouped into three groups of two (K. Adachi et al. 2000). Each pair of subunits can be in one of three states, either empty, bound to NTP or bound to NDP. Each pair of subunits cycles between these three states sequentially. This is difficult to reconcile with the homomeric nature of the ring, in contrast to the F1-ATPase structure in which alternating subunits are of different types.

A fourth model is a sequential model. In this case each of the subunits cycles through states from empty, to NTP bound, through hydrolysis, to release. There may be multiple distinct intermediate states, or only a few. This model is appealing because it agrees with the structural data in that each site is identical, but in progressively different mechanical states. This mechanism does not necessarily have to be sequential around the entire ring, it may be that anywhere between just one, to all six subunits are capable of occupying different states at once. Versions of this model have recently been suggested for T7gp4 (Liao et al. 2005), E1 (Enemark & Joshua-Tor 2006) and Rho (Thomsen & Berger 2009).

These mechanisms are summarized in Figure 3. Hydrolysis states are indicated both by labeling and color of the subunit, progressing from Empty (E); through ATP bound (A); Activated ATP bound (A'), hydrolyzed with diphosphate bound (D.P) and hydrolyzed with diphosphate released (D). These do not necessarily precisely correspond to the physical states.

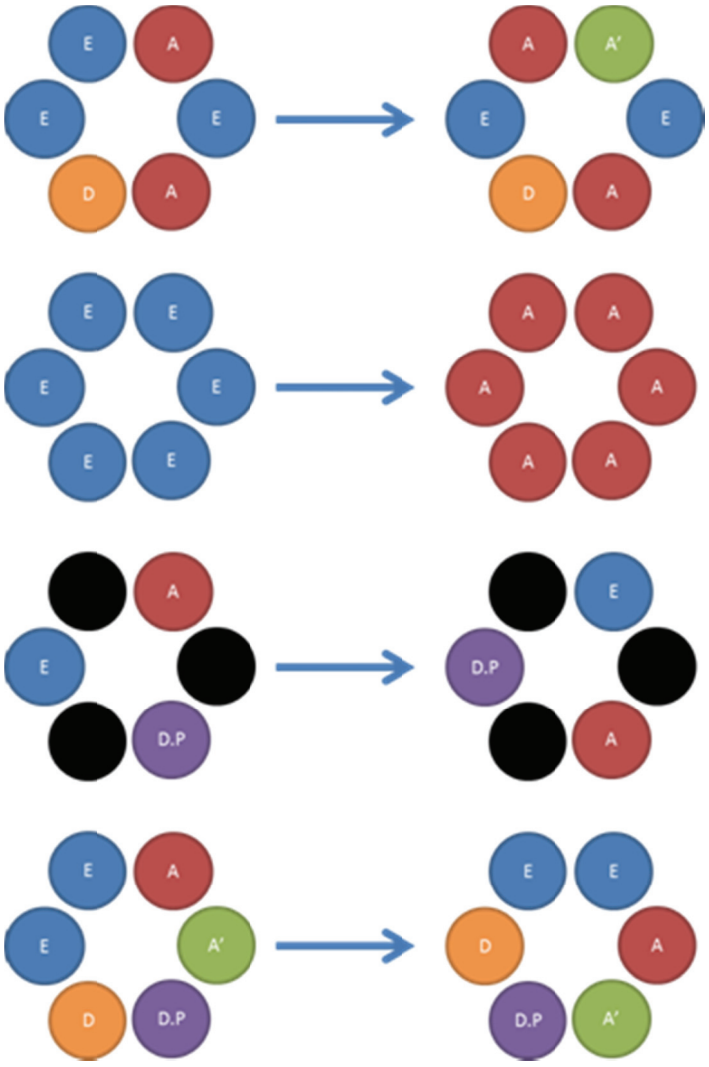

Figure 3) Models for ring-shaped helicase translocation mechanism. From top to bottom: stochastic; simultaneous; grouped sequential (F1-ATPase like); sequential

Experimental approaches to understanding helicase function

There are a variety of experimental approaches to probing the properties and mechanism described above. Each of these approaches have different strengths and weaknesses, and a combination of data from all sources will provide the most valuable and complete set of information. In this section I will examine three broad, common experimental approaches to understanding protein function.

One approach is to solve crystal structures of proteins and to speculate on the function of the revealed structure. This can be challenging for multiple reasons. Firstly, it is not always possible to crystallize a protein, or if it is possible, it may only be possible in very specific conditions which are not the naturally occurring conditions for the protein, such as when complexed with non-physiological analogs. This means that crystal structures may not accurately represent the physiological configuration. Secondly, crystal structures represent only one snapshot of the protein. Since the mechano-chemical cycle is a dynamic process this only gives insight into a small fraction of the many configurations the protein can take. Finally, it is often very difficult to interpret structural features in the context of function. For example, a loop motif may appear to bind to the DNA backbone, but if the DNA and protein are not crystallized together this cannot be verified. The major advantage of this technique is that it provides an exact atomic blueprint of the protein. Without this data it is nearly impossible to fully understand the biological mechanism.

Another experimental approach is to conduct bulk biochemical experiments. The greatest strength of this approach is also the greatest weakness, which is that it is a measurement of the average behavior of an enormously large group of molecules. This is good because enormous amounts of data are being collected by each experimental,

since billions of molecules are being examined at once. It is bad because this makes it hard to probe non-equilibrium parameters, such as the dynamics of a motor protein. For example, this approach has been used to determine the average hydrolysis rate of helicases, and the average translocation rate. However it is nearly impossible to measure the dynamic variation in these rates along a single strand of DNA since the behavior is necessarily averaged over all possible states of the protein on the DNA.

A third approach is to use single-molecule methods to study the behavior of individual proteins on individual strands of DNA. In contrast to bulk studies in which many millions of molecules may be examined and averaged together, single-molecule experiments involve looking at biological molecule one at a time. This has the advantage that an experimenter can watch a protein performing its function in real-time, hence revealing information which would otherwise be lost. For a helicase experiment this means that the unwinding process can potentially be observed as it happens, instead of looking at an equilibrium result. The difficulty with this approach is that statistics are collected extremely slowly, since only one molecule is examined at a time. Experimental design can also be very challenging. Because understanding the mechano-chemical coupling requires an understanding the dynamics of the protein, the single-molecule approach is particularly appropriate.

T7gp4 Helicase Step Size

Each of the above experimental methods has been used to perform many experiments of T7gp4 helicase. However, one piece of data which has so far eluded experimenters has been a measurement of the physical step size. In fact, the physical step-size has never been observed for any hexameric helicase, although the evidence supporting a sequential or partially sequential model described above strongly suggests that it is

likely to be 1-2bp. Direct observation of a step-size in this range would be an important confirmation which ties together the structural, biochemical and existing single-molecule data.

In order to observe the step-size, it is necessary to be able resolve 1nm displacements. This is because the contour length of ssDNA is around 0.5nm, and it only reaches this contour length at around 20pN, meaning that under reasonable experimental conditions it will be even smaller.

This thesis describes the construction of an instrument capable of resolving these distances (Chapter 2), simulations and theoretical work to determine the magnitude of the signal as well as the background noise (Chapter 3), the synthesis of an appropriate DNA construct to use with the instrument, as well as experimental results (Chapter 4), and finally a discussion of relevant extensions to this work (Chapter 5).

Summary

Helicases are ubiquitous proteins which convert dsDNA into ssDNA in a variety of diverse biological processes. They share motifs with many other translocases and ATPases, and by understanding the mechanism of a single protein we can better understand the entire class. We study T7gp4, a ring-shaped helicase from bacteriophage T7. Current experiments indicate that the physical step-size of this protein is likely to be 1-2bp, but this has never been directly observed.

Chapter 2:
Construction of a Dual Optical Trap

Introduction

In order to resolve 1 basepair unwinding events, we must be able to resolve changes in DNA length on the order of 1 nanometer. Optical trapping is an ideal technique to use due to the high bandwidth and resolution afforded by back focal plane detection. However, the standard configuration for performing such experiments in a single trap is unsuitable due to significant instabilities at small length scales on experimental timescales. In order to observe these displacements a more sensitive instrument is required. A dual-trap instrument has better sensitivity and is immune to the drift issues which affect single trap instruments.

Optical Trapping

Optical trapping is a technique for manipulating microscopic particles using a highly focused beam of light (Ashkin 1970; Ashkin et al. 1986). Dielectric particles in the presence of the field produced by the light are pulled to the region of highest intensity by gradient forces (Ashkin et al. 1986). The particle will be displaced slightly from the center in the vertical direction due to the additional force of radiation pressure. The position of the particle in the trap can be determined at very high speeds using back focal plane detection (Keir C. Neuman & Steven M. Block 2004). By calibrating the stiffness of the trap, and the sensitivity of the detection system the force and displacement of the particle can be measured at rates limited only by the detector and data acquisition speed.

Standard Optical Trapping Configuration

A simplified layout for a standard single optical trap is shown in Figure 4. Light from a 1064nm TEM00 continuous wave laser is first passed through an acousto-optical modulator (AOM), and the zeroth order beam is blocked. The AOM allows the

proportion of energy in the first order beam to be electronically controlled, thereby permitting the intensity and strength of the trap to be modulated (Korpel 1996). It also acts to frequency-shift the beam, so that any light reflecting back into the laser cavity will exhibit minimal interference and mode-hopping.

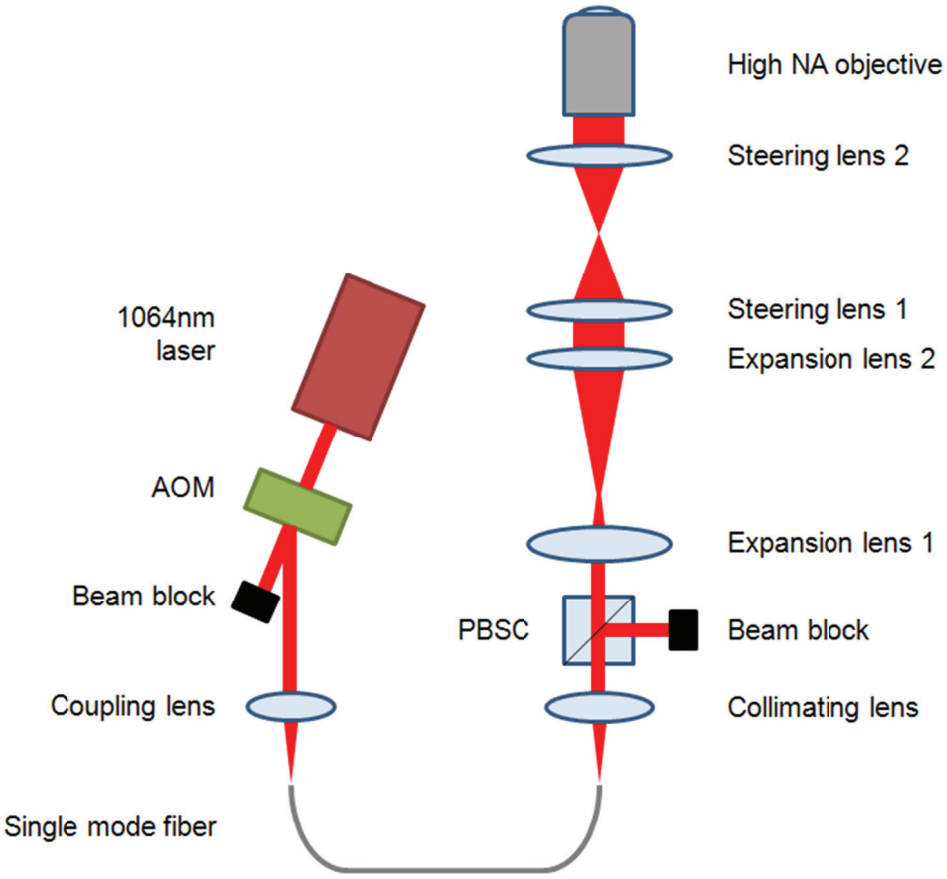

Figure 4) Schematic layout of single beam optical trap

The output from the AOM is then coupled into a single-mode optical fiber. This converts pointing instabilities of the laser into power fluctuations. Power instabilities are preferred to pointing instabilities since the latter results in an increase in positional

noise of the beam, whereas the former can be corrected for by performing all experiments using position measurements which are normalized to the total incident power. Positional changes of the beam in the sample chamber are indistinguishable from motions of the trapped particle. By coupling the laser to a single mode fiber all pointing instabilities simply correspond to a change in coupling efficiency, changing the total power output of the fiber.

The output from the single mode fiber is then collimated and the polarization is cleaned using a polarizing beam splitting cube (PBSC). This leaves a single, pure polarization component. This will also convert any polarization noise into power fluctuations. At this point, the polarization can be arbitrarily rotated using a $\lambda/2$ wave-plate (Hecht 2001). This may be important if using a microscopy configuration which relies on polarization.

Next, the beam is expanded. This is so that the waist of the beam overfills the back aperture of the microscope objective, resulting in a stiffer trap then would otherwise be obtained (Fallman & Axner 2003). This is accomplished with a pair of plano-convex lenses with the ratio of focal lengths determining the magnification. Typically the beam is expanded so that the $1/e^2$ points of the Gaussian beam match the objective aperture (Keir C. Neuman & Steven M. Block 2004).

The beam may also be passed through an additional 1:1 telescope, where the initial lens is mapped to the back aperture of the microscope objective. This telescope does not affect the structure or size of the beam, but rather serves to link displacements of the lens to displacements of the beam in the sample plane. This occurs because when correctly mapped, translations of the primary lens will correspond to a pure rotation at

the back aperture of the objective lens (Hecht 2001). This allows manual control over the beam position in the sample plane. By mounting the lens on a micrometer stage an experimenter can control the beam position by hand with a resolution of tens of microns. This is useful for large adjustments before conducting an experiment.

Finally the beam passes through the objective lens of the microscope. This is typically a high numerical-aperture, oil or water immersion lens. A high NA lens is vital to produce a strong gradient, giving a large stiffness (Keir C. Neuman & Steven M. Block 2004). A low NA lens will result in a trap with an axial stiffness which is unable to compete with the scattering force, and will not trap particles. The diffraction limited spot formed by the focused beam forms the optical trap.

Back Aperture Detection

Detection is accomplished by collecting the light exiting the condenser lens. Typically this is separated from the illumination path of the microscope using a dichroic mirror which reflects the infrared trapping beam, but does not affect the visible illumination source for the microscope. The light forming the optical trap passes through the sample chamber and exits through the condenser aperture.

The light is then collected by a position sensitive detector (PSD). The plane of the detector is mapped onto the back aperture of the condenser using a telescope. This mapping provides the greatest sensitivity to deflections of the beam. If the beam is mapped correctly displacements of a trapped bead will correspond directly to a shift in position of the beam at the detector (Gittes & Schmidt 1998). A typical detector of choice is a quadrant photodiode, in which the power falling on each of four quadrants of a circle is used to perform a weighted spatial average.

The detectors are adjusted so that when a free bead is trapped in the beam, the position of the beam on the detector is zeroed in both dimensions. As the bead is displaced in the X or Y directions, the beam is also displaced at the back aperture of the condenser lens (Gittes & Schmidt 1998). Since the PSD is mapped to the back aperture, the displacement of the bead results in a corresponding change in position of the beam on the detector. By calibrating the displacement of the bead to measurements at the PSD, the position of a trapped bead can be determined to sub-nanometer resolution, at rates limited by the bandwidth of the detectors and the data acquisition rate, which is typically at least 10KHz.

The force on a displaced bead can also be determined from the position, if the trap stiffness is known. The TEM00 mode has a Gaussian intensity profile, and the force is proportional to the first derivative of the intensity (Ashkin et al. 1986). This means that in the center of the trap the force scales linearly with displacement, and also with laser power. The stiffness can be calibrated using a variety of methods (Keir C. Neuman & Steven M. Block 2004; Tolić-Norrelykke et al. 2006), and once determined, the position measurement combined with a measurement of the total beam intensity can be converted into a force. A typical trap stiffness is 1pN/nm.W, at typical powers this means that forces of up to 100pN can easily be generated on the bead.

Once calibrated, position data from the PSD's along with a measurement of the beam power in the sample plane can determine the position and force of the bead in real-time, at high accuracy and precision. If the stage position and trap height are known this can be converted into information about the extension of the DNA molecule being probed.

Additionally, the power and extension can be modulated by feeding back to the AOM and the piezo-electric stage. This allows an experimenter to perform a variety of "clamps" to probe different aspects of a biological system.

To stretch a molecule at a certain rate a velocity clamp can is used. In this mode the position of the bead is monitored while the stage is moved at a constant rate. If the bead is displaced from its set-point in the trap then the intensity of the beam is modulated to either increase or decrease the trap stiffness, so that the bead is pulled back to the desired location. This keeps the bead at a constant location in the trap, resulting in a constant velocity of extension of the DNA construct.

Another common clamping technique is to use a force clamp. In this configuration the position of the bead is also maintained by modulating the beam intensity, but instead of moving the stage at a constant rate, the stage is displaced so as to maintain a constant force on the bead. This is useful for experiments in which a protein is causing a change in the number of ssDNA bases, such as the helicase unwinding assay.

A third clamp is a force-ramp, which is similar to the force clamp but with a varying force set-point. This is useful for dynamic force spectroscopy type experiments in which the loading rate is an important parameter (Strunz et al. 1999).

There are a variety of more complicated clamps, but generally these are extensions and combinations of the three basic clamps described here. In general the feedback is performed by using PID loops running at rates close to 1kHz.

Experimental configuration in single trap

A typical experimental configuration for a single trap is shown in Figure 4.

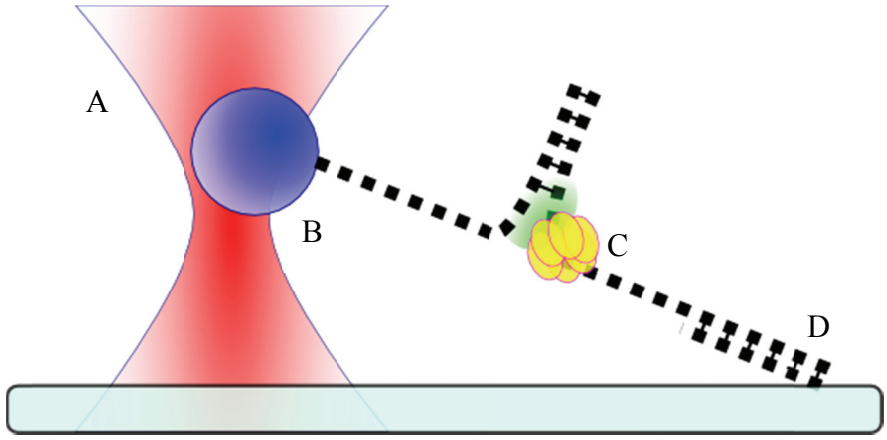

Figure 5) Typical configuration for helicase unwinding experiments. A) Optical trap; B) Polystyrene bead; C) T7 Helicase; D) Double-Stranded anchoring segment

In single trap unzipping experiments one end of the DNA construct is attached to the surface of a cover-slip using a digoxigenin/anti-digoxegenin antibody interaction via a modified base. The other end of the construct is attached to a 500nm diameter, coated polystyrene bead using a streptavidin/biotin interaction (D. L. Nelson & Cox 2004). This bead is then trapped. The tension in the DNA can be monitored via the measured force on the bead, and controlled by translating the coverslip in any spatial dimension using a piezoelectric stage. By combining knowledge of the end-to-end DNA extension and the tension in the DNA, the number of single-stranded basepairs can be determined.

The DNA can be unzipped by applying a force with the stage, or it can be unwound by a protein such as T7gp4 helicase. In this case the force can be held at a specific value

as the protein unwinds the DNA by feeding back on the stage position as progressively more basepairs are released. This provides an assisting force to the helicase.

For example, by performing a force clamp after adding helicase the position of the stage will be steadily increased as more double stranded basepairs are unwound into the single stranded arms. This is because as bases are added to the arms the total stiffness decreases, so in order to maintain the force set-point the stage must be displaced, extending the DNA in the arms. In this manner it is possible to assist the helicases unwinding activity by providing a force at the junction which acts to destabilize the dsDNA bases.

By varying the assisting force and measuring the rate at which the helicase unwound the DNA we were able to determine the activeness of the enzyme, as well as predictions for the step-size (D. S. Johnson et al. 2007).

Drift in single optical trap configurations

Noise can be a significant problem in optical trapping experiments. One of the greatest sources of noise is due to the Brownian fluctuations of the bead in the trap, the characteristics of which depend on parameters including bead size, DNA stiffness, and temperature (Keir C. Neuman & Steven M. Block 2004). This noise imposes a basic limitation on the resolution of the instrument. It can be reduced by careful choice of experimental design -- for example by minimizing the length of anchoring DNA -- but it can never be eliminated.

Instrumentation noise can also be reduced. This noise has contributions due to electronic noise, vibrations coupled into the microscope, ground loops, and various

other sources. One of the strongest sources of noise which affects single trap instruments is fluctuations in the position of the cover-slip relative to the optical trap. This is generally referred to as microscope drift. A typical drift rate is on the order of a nanometer per second in the axial direction, and it can be in either a positive or a negative direction. Possible causes of this drift component are the viscous immersion oil settling and dragging on the coverslip, liquid evaporating from the sample chamber, and local heating effects. Although it is possible to stabilize this drift to within a nanometer over a long period of time (A. R. Carter et al. 2007), deviations on short time scales are hard to accurately correct for, and it is exactly these which are likely to obscure individual unwinding steps.

I performed experiments to measure typical amounts of drift on two separate instruments. Representative data is shown in Figure 6. Drift typically exhibits both long term components -- such as the long, slow downward trend – as well as faster components. The drift in the plane perpendicular to the optical axis of the microscope is orders of magnitude less than that in the axial direction, and is typically not large enough to affect experimental data.

The drift was tracked by performing image correlation on beads fixed to the surface of the sample chamber. This provides a resolution of < 5nm. Since this only tracks the physical motion of the coverslip it does not include additional drift contributions due to trap position fluctuations, such as those due to heating of the objective lens (Mahamdeh & Schaffer 2009). Measurements which track the position of the bead on the coverslip will thus always be an underestimation of the total drift which occurs during a trapping experiment.

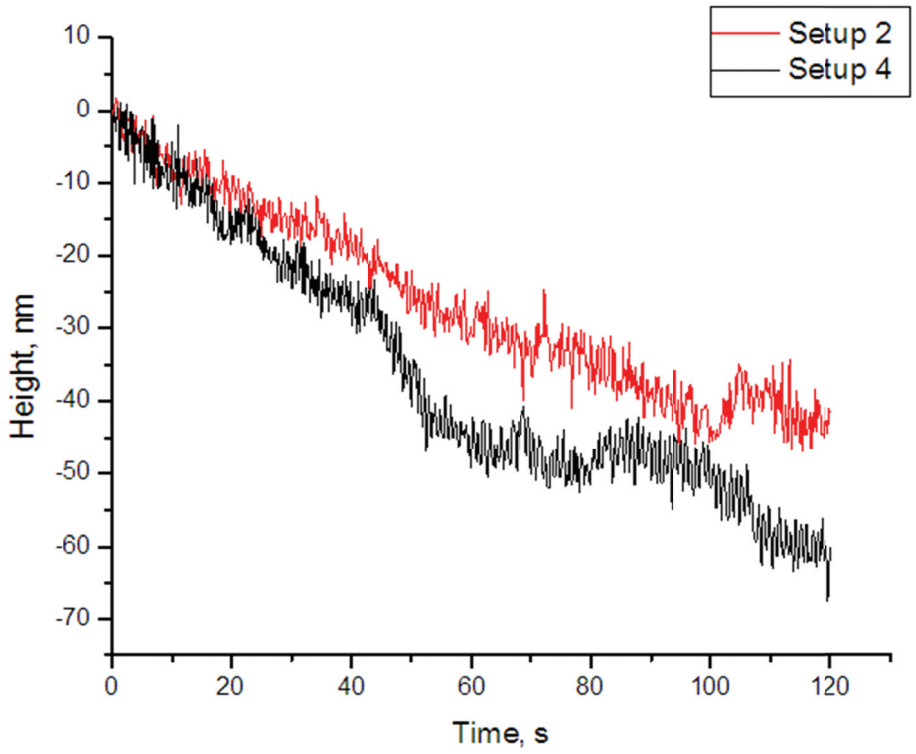

Figure 6) Drift on two optical trapping instruments. The vertical position of a 479nm polystyrene bead stuck to the cover of a coverslip was tracked for 2 minutes on two of our optical trapping instruments.

One way to correct for drift is to feed-back on the stage position by using a stuck bead or other fiducial mark (Nugent-Glandorf & Perkins 2004). By monitoring the position of this mark, changes in the position of the stage can be countered by raising or lowering a piezo stage on which the sample chamber is mounted. I implemented this by performing feedback using video tracking of a bead stuck to the surface of the sample chamber.

An example of this is shown in Figure 7. In the data shown the stage was driven in 1nm steps with the piezo stage while the position of a stuck bead was monitored using back focal plane detection. At t=85s drift correction was activated while the stage continued to be stepped. As can be seen, this method can stabilize the stage to <1nm over very long periods of time, but on shorter time scales there are still deviations of several nanometers.

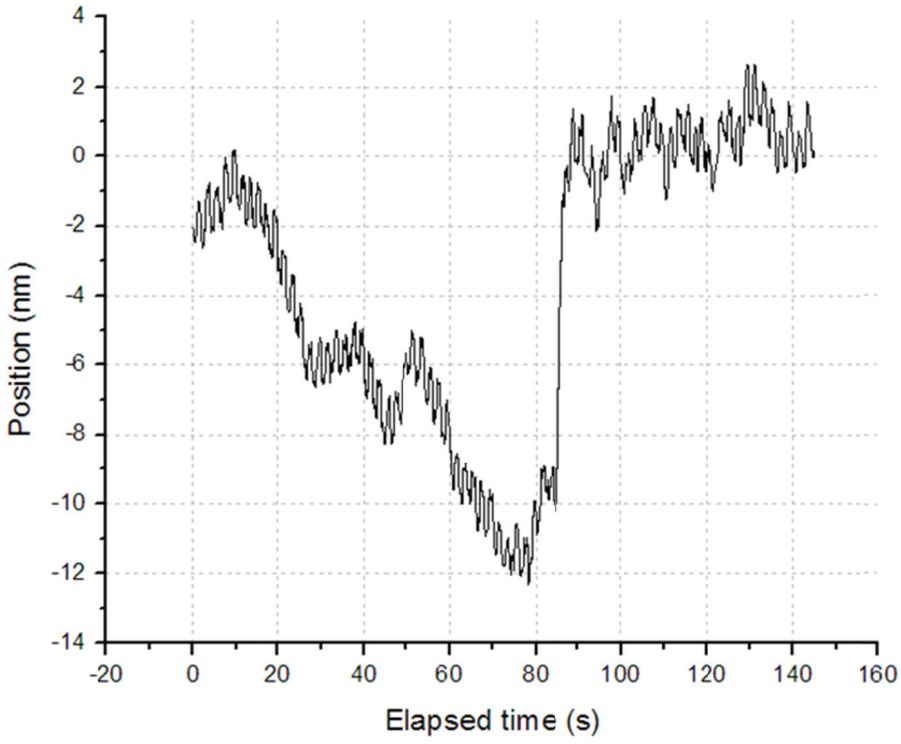

Figure 7) Drift correction in Z of a microscope stage. The piezo stage is driven in a 1nm square wave, and the position is monitored using back-focal plane detection. At t=85s drift correction is activated.

The drift in coverslip position affects the calculated number of unzipped basepairs. This is because the DNA is attached to the surface at an angle, so that the height at

which the bead is held relative to the surface is required to correctly calculate the extension of the DNA. If this height varies due to variations in the height of the coverslip, this will be added to calculations of the DNA extension. The short-term fluctuations are of particular concern for step size measurements because they are on the order of the signal generated by a few basepairs being unwound.

The drift in coverslip position also affects the stiffness of the trap. Our previous experiments have all been carried out using oil immersion lenses. Because of the index of refraction mismatch between the liquid in the sample chamber (n=1.33) and the immersion oil (n=1.51) the degree of spherical aberration depends upon the height at which the beam is focused relative to the coverslip (Pawley 1995). This manifests as a change in trapping stiffness as a function of trapping height, and also means that at distances greater than around 10um in the chamber it is very hard to trap beads.

Lastly, any other uncorrected fluctuations in the quality of the beam, such as deviations in position due to movement of optical components, or Schlieren noise due to index of refraction fluctuations of the air will all add noise to the measurement.

Dual Trap

A dual trap configuration eliminates or reduces many sources of noise. Instead of holding a tether between a trap and the surface of the coverslip, a second trap is formed, and a second bead attached to the end of the DNA construct. This completely eliminates any noise from stage drift by completely decoupling the experiment from the surface. Any relative changes in the trap position with respect to the chamber surface will be experienced equally by both beams, and so the relative heights will remain the same.

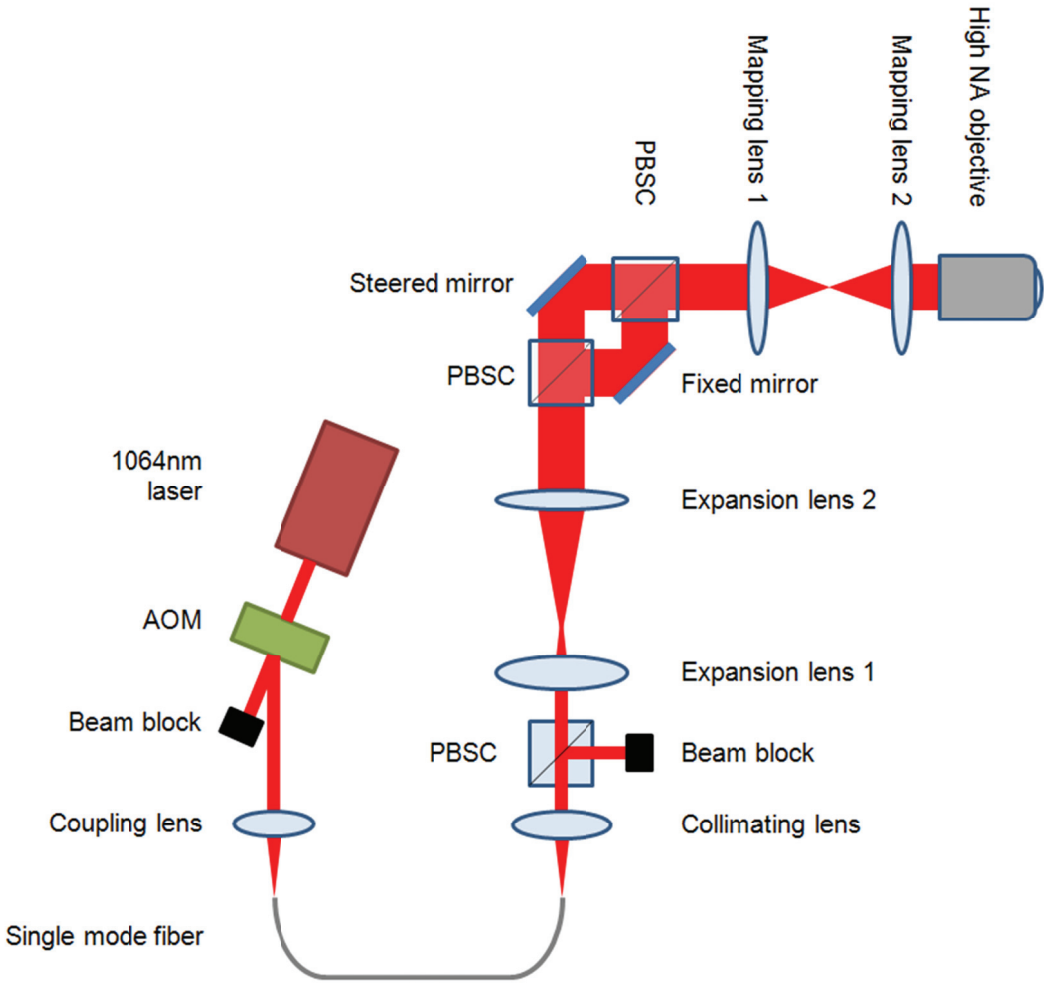

Figure 8) Schematic layout of dual optical trap

A schematic layout is shown in Figure 8. This resembles the single trap configuration, but with an additional section before the final telescope, in which the beam is split based on polarization. This is accomplished by using polarizing beam-splitting cubes (PBSC). One beam then passes through a "standard" optical path, whereas the second reflects off of a piezo mounted steerable mirror, which is mapped to the back focal plane of the objective lens. The beams are then recombined with a second PBSC. By

adjusting the polarization of the incoming beam the relative intensities of the two beams can be continuously altered.

By rotating the steerable mirror the second trap can now be translated in the sample plane. Because the mirror is conjugate to the back aperture of the objective lens, a rotation of the mirror corresponds to a pure translation of the trap in the sample plane. The second trap now plays an analogous role to the stage in the single trap setup. By adjusting the position of this beam relative to the fixed beam, forces can be applied to the DNA. The position of this beam can be adjusted based on the measurements acquired by the acquisition system.

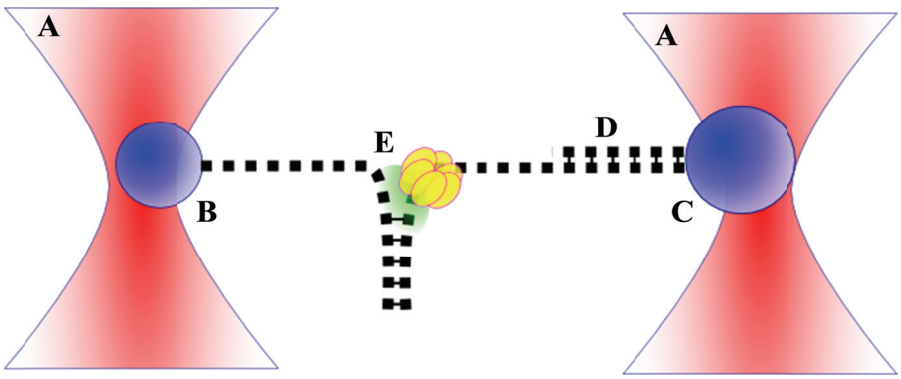

Figure 9) Dual-beam optical trap configuration. A) Optical traps; B) Streptavidin coated bead; C) Anti-Digoxegenin coated bead; D) dsDNA anchor region; E) T7 Helicase, green region shows destabilization interaction.

Fluctuations in the cover-slip position now affect both beams equally. Previously, if the height of the trap changed due to fluctuation of the coverslip position, this added noise to the calculation of tether length. In the dual trap configuration both traps are

shifted by the same amount, and so fluctuations in the height of the coverslip do not affect the extension of the DNA.

Additionally, any pointing instability of the beam which occurs before the beams are split will also affect both beams. To maximize the advantage that this provides the optical path in our instrument is designed so that the split path portion is as close to the objective lens as possible, and goes through as few optical elements as is possible. This means that there are only a few pieces of optics which can affect the beams in a non-identical manner.

Water immersion objective lens

Using a water immersion lens is also beneficial for reducing the effects of noise. With an oil immersion lens the absolute focal position does not stay constant when the sample chamber is displaced. This is shown in the simulation in Figure 10a. This means that when performing experiments with an oil immersion lens we must apply a correction for the relative motion of the trapping position as the stage is moved in the Z direction. With a water immersion lens the trapping position remains at the same absolute position, shown in Figure 10b.

The degree of spherical aberration increases with trapping depth in an oil immersion lens. This means that the trap stiffness varies with height, and it is not possible to reliably trap deeper than around 10um into the sample chamber. The amount of spherical aberration when using a water immersion lens only depends upon the thickness of the coverslip. This can be eliminated for a certain thickness by adjusting the correction collar located on the body of the lens.

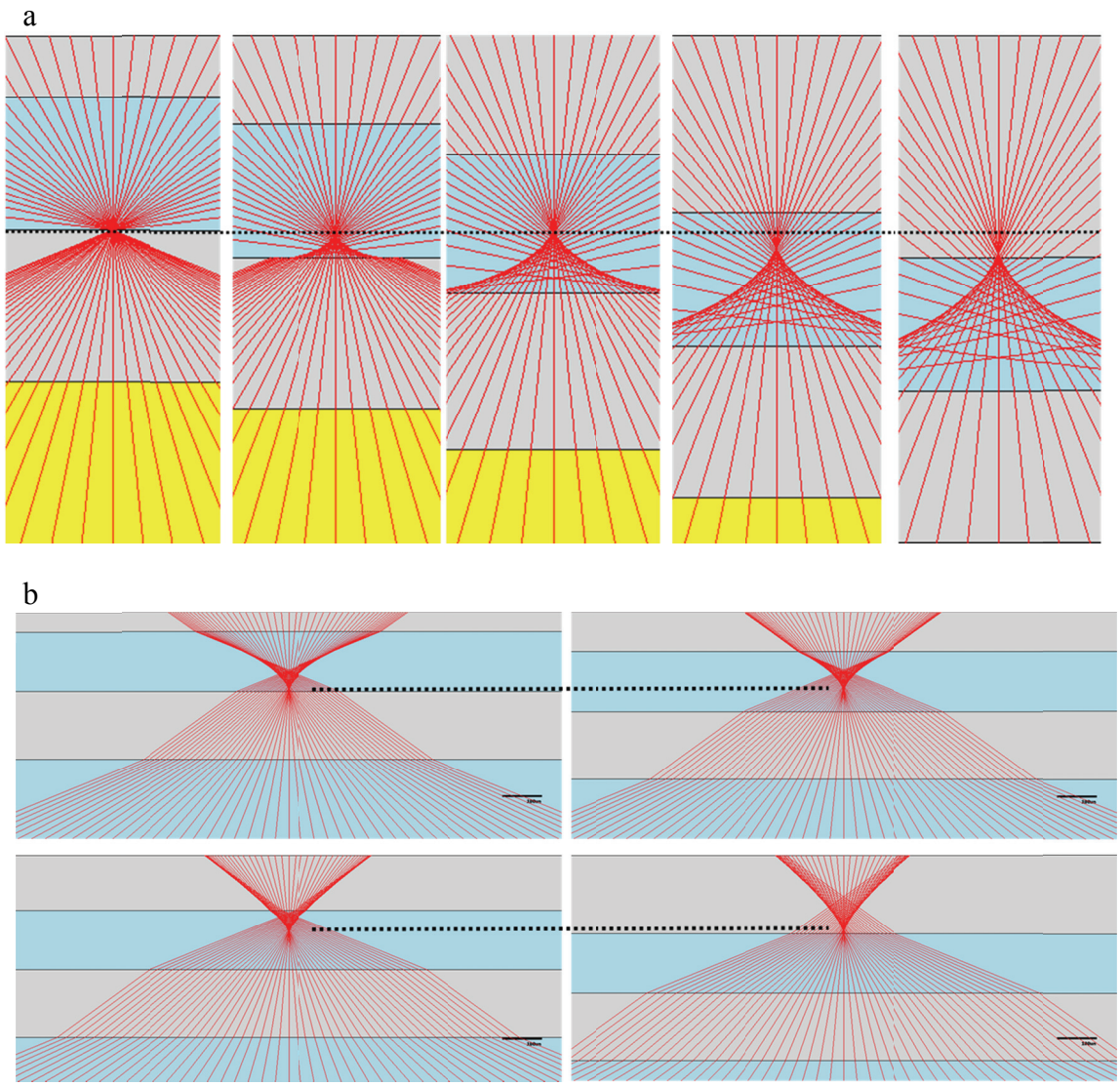

Figure 10) Comparison of the importance of height for a) oil immersion and b) water immersion optical trapping. In oil the absolute position of the beam focus shifts proportionally to the height of the coverslip, and there is greater spherical aberration at greater depths. In oil both position and degree of aberration are constant.

When using a dual-trap configuration in conjunction with a water immersion lens there is no change in trap stiffness or DNA extension with coverslip height. It is thus

an ideal configuration for observing small signals such as the unwinding steps of T7 helicase.

Differential Measurements

In addition to isolating the DNA construct from the coverslip, there is an additional important advantage which a dual trap provides over a single trap configuration. It allows one to make differential measurements of position, by using the difference co-ordinate of the beads rather than the displacement of just one bead. This is advantageous because Brownian fluctuations will tend to move the beads in the same direction, since they are coupled together by the DNA. However, an unwinding event will cause them to move in opposite directions (Moffitt et al. 2006).

Our Instrument

The dual trap I built used a 5W YVO_4 continuous wave 1064nm infra-red laser (Spectra-Physics). This passed through an AOM (Isomet) and was then coupled to a single-mode fiber (OZ Optics). The output was passed through a $\lambda/2$ waveplate, a PBSC, a second $\lambda/2$ waveplate before entering the primary expansion telescope. Next the path is split with a PBSC, one path is reflected from a fixed dichroic mirror and the second from a mirror mounted on a tip-tilt piezo stage (Mad City Labs). The paths are recombined and passed through a combination mapping/expansion telescope before entering a modified microscope body (Nikon). The objective used is a 1.2NA 60X water immersion lens (Nikon). The output beam passes through a mapping telescope, is split by a PBSC and each beam is incident on a quadrant photodiode (Pacific Silicon Detectors). The signal is filtered by an anti-aliasing 8-pole Butterworth filter (Krohn-Hite) and acquired by a 16bit data acquisition card (National Instruments). All

feedback, acquisition, tracking and analysis software was custom written in LabView (National Instruments).

Examples of dual trap resolution

An example of the immunity from drift is shown in Figure 11. This shows how the calculated number of basepairs fluctuates as an unzipping construct is held open, for a single trap and for a dual trap configuration. Because the calculated number of basepairs depends on the extension of the DNA, drift will affect this calculation for the single trap since it affects the. This can be seen in the top trace in Figure 11. Drift does not significantly affect the dual trap because the traps are isolated from the surface. This is shown in the bottom trace.

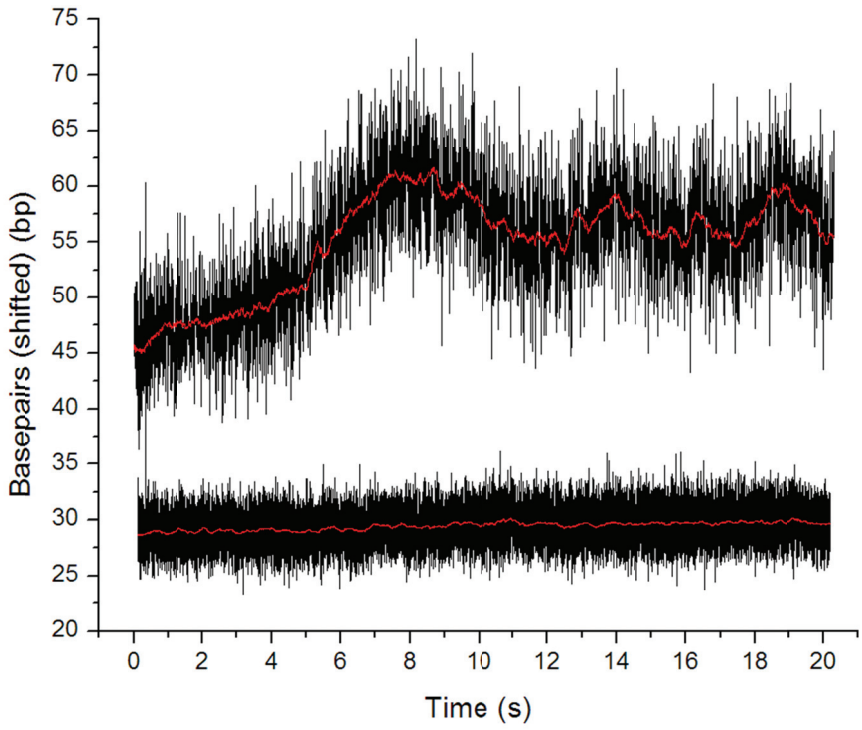

Figure 11) Comparison of trapping stability for single trap (top) and dual trap (bottom).

Figure 12 shows an example of the resolution of the dual-trap instrument. This shows a 3kbp dsDNA molecule stretched between the traps. Every second the movable trap is stepped by 2nm, resulting in a change of extension slightly greater than 1nm.

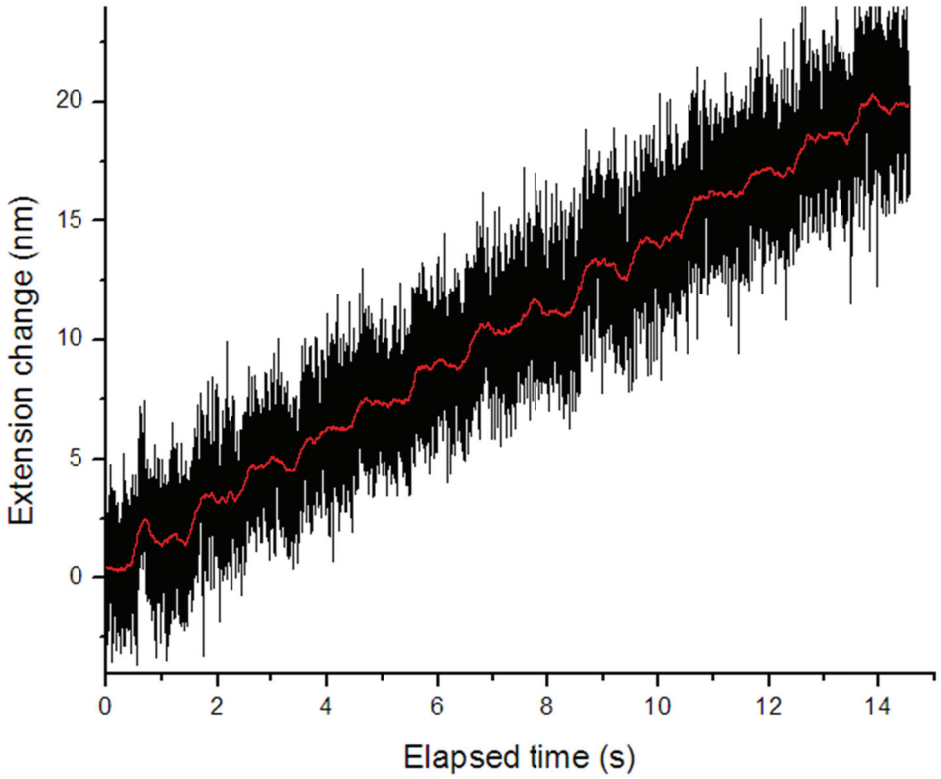

Figure 12) Stretching a dsDNA in steps, 2000Hz data in black, adjacent averaged to 40Hz in red.

Summary

Single beam optical traps are the most common type of optical trapping apparatus. They are capable of resolving sub-nanometer motions of a trapped bead, and forces to sub-piconewton precision. When performing experiments with a single beam optical trap the DNA construct must be tethered to the surface of the sample chamber. This

couples in noise due to drift in the stage position which occurs at rates of up to 1nm/s. When performing unwinding experiments this reduces the accuracy and stability of the base-pair measurement. For high-precision experiments this drift can obscure the signal of interest. To decouple the experiment from stage motions a dual optical trap can be used. This uses a second bead and beam, instead of attaching the DNA to the coverslip. In this way the noise from stage drift is removed, and additionally the differential signal improves resolution. The dual trap instrument I built shows a significant improvement in stability and small increase in resolution compared to our single-beam instruments.

Chapter 3:
Simulations of the Helicase Unwinding Signal

Introduction

To examine whether the advantages of the dual-trap afford us the necessary resolution to observe individual unwinding events, we must identify the magnitude of the expected step signal and the magnitude of the noise signal over the measurement bandwidth.

In order to do this we must know both the expected signal from a single physical step, and the background noise from the fluctuations in the system due to Brownian motion. These parameters may be estimated from equilibrium simulations of the system. In the following chapter the method of simulation will be described, as well as the predictions resulting from this work.

Force Extension Relations in DNA

The first step in this procedure is to be able to accurately and quickly calculate the force-extension curves of ssDNA and dsDNA. These non-linear curves have been well characterized by theory and experiment (S. B. Smith et al. 1996; M D Wang et al. 1997). Examples of force-extension curves for both ssDNA and dsDNA are shown in Figure 13.

Both ssDNA and dsDNA exhibit a non-linear, entropic region below their contour length (0.34nm for dsDNA, 0.54nm for ssDNA) followed by a rise to a more linear regime when the contour length is exceeded. The approximations used in this work neglect the more complicated effects which occur at higher forces. For example, the dsDNA overstretching transition which occurs at around 60pN does not appear on the figure below (S. B. Smith et al. 1996). This is not important for these calculations as

the forces used in experiments are always below the critical unzipping force, which is never more than 20pN.

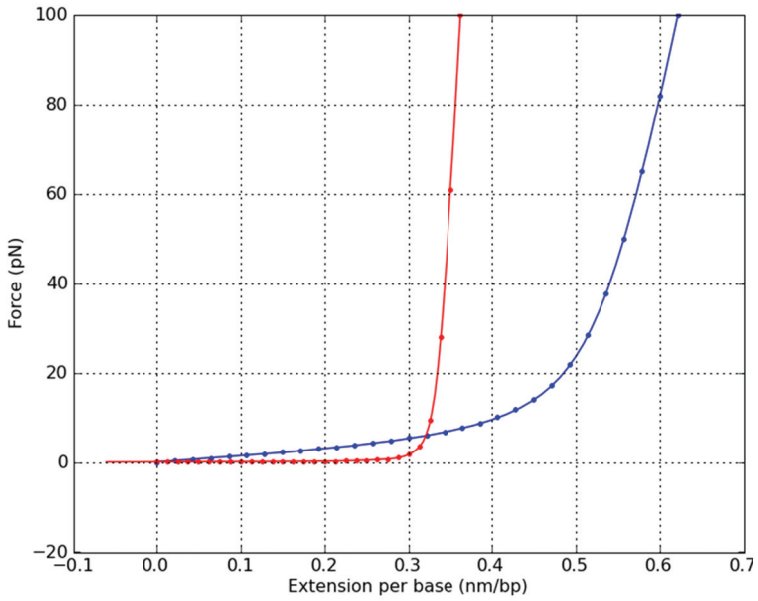

Figure 13) Force-extension curves for ssDNA (blue) and dsDNA (red). Points show exactly calculated values, whereas the lines show spline interpolations to the curve, using 50 points between 0 and 50pN.

ssDNA is well fit by a modified freely jointed chain model, where the extension-force relation is given by (S. B. Smith et al. 1996):

$$x(F) = L_{ss}\left[\coth\left(\frac{Fb_{ss}}{K_bT}\right) - \frac{K_bT}{Fb_{ss}}\right]\left(1 + \frac{F}{K_{ss}}\right) \quad (1)$$

Where L_{ss}=0.537nm is the single-stranded contour length, K_b is Boltzmann's constant, T is temperature, b_{ss} =1.59nm is the single-stranded Kuhn length (twice the persistence length), and K_{ss}=533.3 pN is the single-stranded stretch modulus.

There are several good approximations for the dsDNA force-extension relation (M D Wang et al. 1997). In this work I use the modified Marko-Siggia wormlike chain model for forces up to 20pN:

$$F(x) = \left(\frac{2K_bT}{b_{ds}}\right)\left[\frac{1}{4\left(1 - x/L_{ds} + F/K_{ds}\right)^2} - \frac{1}{4} + \frac{x}{L_{ds}} - \frac{F}{K_{ds}}\right] \tag{2}$$

And I use the Odijk model at forces above 20pN (Odijk 1995):

$$x(F) = L_{ds}\left[1 - \frac{1}{2}\left(\frac{2K_bT}{Fb_{ds}}\right)^{\frac{1}{2}} + \frac{F}{K_{ds}}\right] \tag{3}$$

Where L_{ds} =0.34nm is the double-stranded contour length, K_b is Boltzmann's constant, T is temperature, b_{ds} =86.2nm is the double-stranded Kuhn length, and K_{ds} = 1205 pN is the double-stranded stretch modulus.

In order to increase the speed of calculation, the force extension curves for both ssDNA and dsDNA are pre-calculated and stored in lookup tables. Values are subsequently obtained by performing a spline interpolation between the calculated points. In Figure 13 above the points show exactly calculated values and the curves are obtained by interpolation.

To understand the calculation of the unzipping signal magnitude it is helpful to first consider the simpler process of mechanically unzipping the dsDNA by applying a force with the trap. This process may either be performed as a constant-extension experiment in which the separation of the traps is increased in well-defined steps and

the force is allowed to fluctuate, or a constant force experiment in which the force is increased in well-defined steps and the extension is allowed to fluctuate.

Constant Extension

In a constant extension configuration the total extension of the ssDNA is steadily increased, while the force is adjusted via feedback, by modulating the intensity of the trapping beam.

At every possible extension there are a variety of configurations the DNA can take. There could be relatively few basepairs in the ssDNA region, each significantly stretched out. Alternatively there could be a large number of ssDNA bases, each with only a very small extension. Each possible configuration has a certain free energy, with contributions from the elastic energy of the ssDNA, and from the sequence-dependent base-pairing energies.

$$\Delta G(n,x) = \sum_{i=0}^{n} \Delta G_{bp}(i) + 2n\int_0^x F(x')dx' \tag{4}$$

The elastic energy of the ssDNA may be calculated using the previously defined expressions. For constructs which contain a dsDNA anchor region an additional energy term can added. In order to include the dsDNA anchoring segment we add an additional stretch energy term:

$$\Delta G(n_{ss}, n_{ds}, x) = \sum_{i=0}^{n_{ss}} \Delta G_{bp}(i) + 2n_{ss}\int_0^{x_{ss}} F_{ss}(x')dx' + n_{ds}\int_0^{x_{ds}} F_{ds}(y')dy' \tag{5}$$

And the single-stranded and double-stranded extensions can be found for a specific number of basepairs by optimizing the force until the sum of the extensions is equal to the specified total extension.

The precise base-pairing energy depends upon the particular buffer conditions used but is typically around 3.4KbT for GC bases (with 3 hydrogen bonds) and 1.4KbT for AT bases (with 2 hydrogen bonds) (D. L. Nelson & Cox 2004). This means that the base-pairing energy is highly sequence dependent. For regions in which there is high GC content there will be a significantly higher free energy associated with base-pairing than in a region with primarily AT basepairs. A more complete treatment of base-pairing energy also includes "stacking" interactions between adjacent bases. This results in 16 possible different base-pairing energies depending on both the current and subsequent base. As an increasing number of basepairs are broken, the total basepairing free energy monotonically increases, with variations away from being purely linear due to the sequence.

To calculate the equilibrium force for a specific extension the total free energy for each possible number of unzipped basepairs is calculated by adding the contributions from each of terms in the above equation.

Once we have calculated the free energy for every possible number of unzipped basepairs, we can take an average using the calculated free energies to weight each configuration:

$$P(m,x) = \frac{\exp\left(-\Delta G(m,x)/K_b T\right)}{\sum_{i=0}^{N} \exp\left(-\Delta G(i,x)/K_b T\right)} \tag{6}$$

$$\langle m \rangle = \sum_{i=0}^{N} iP(i,x) \tag{7}$$

Where N is the total number of basepairs in the construct. An example of these calculations is shown in Figure 14 below.

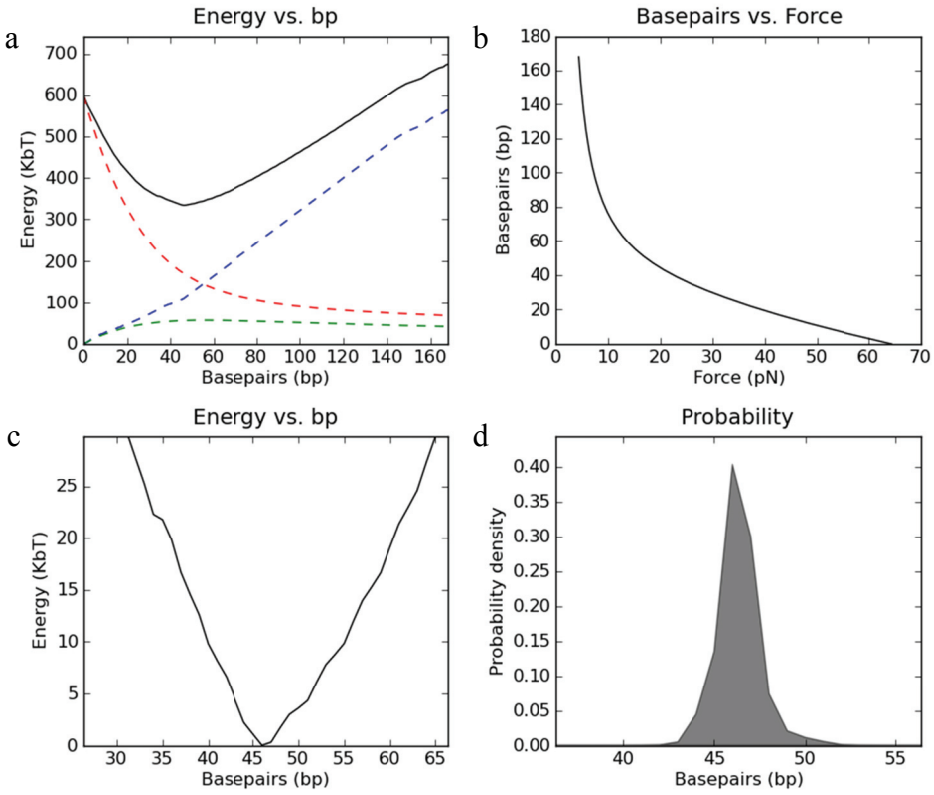

Figure 14) Calculation of basepairs unzipped for a non-homogeneous DNA sequence. a) ssDNA (red), dsDNA (green), sequence (blue) and total (black) free energy contributions; b) basepairs-force relationship; c) enlarged total energy curve; d) probability density vs basepairs

Figure 14a shows each of the three energy contributions separately, and their sum. It is clear from the enlarged sum curve shown in Figure 14c -- in which the energy has

been shifted so that the minimum occurs at 0 for clarity -- that the minimum free energy occurs when 46bp have been opened. This directly results in the maximum probability occurring at the same base, as shown in Figure 14d.

For other sequences the potential energy and hence probability may take a more complicated shape, as shown in Figure 15. Here the equilibrium position will not occur at any of the most likely base's peaks, but in between.

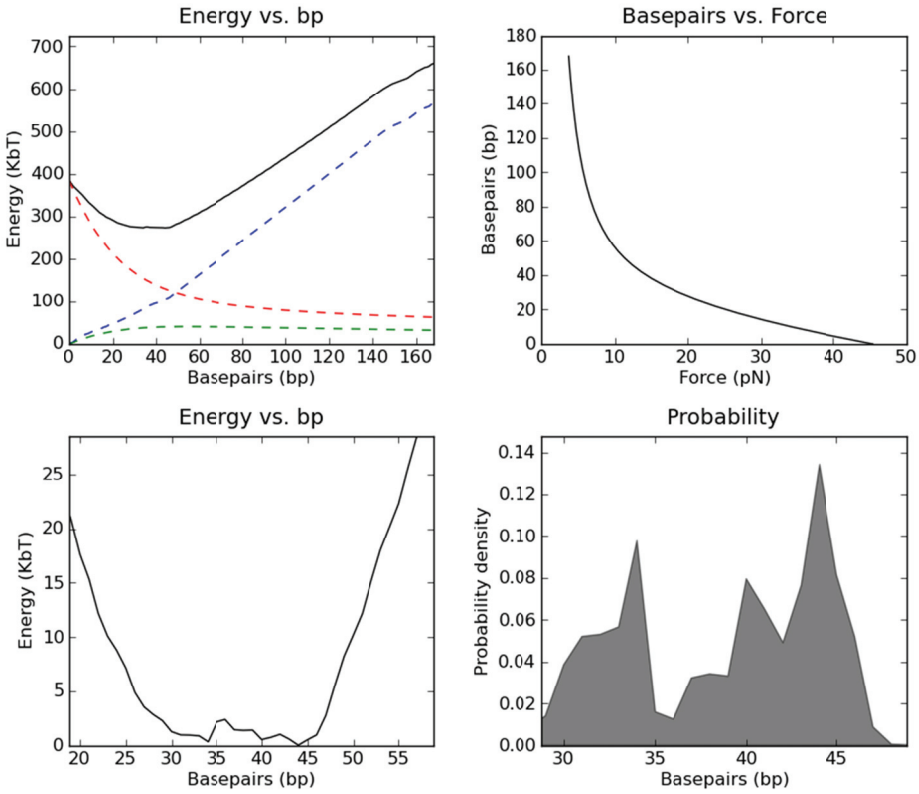

Figure 15) Energy and probability showing more complicated strcuture. Plots are the same as in Figure 14

Because the equilibrium force changes as a function of extension, this produces a characteristic stick-slip shape for the force-bp curve, such as that shown in Figure 16.

This is due to the sequence dependent contribution of the base-pairing energies. In AT rich regions the force is low. In GC rich regions the force is high. However, the typical unzipping force is around 15pN, with variations of around 3pN in either direction.

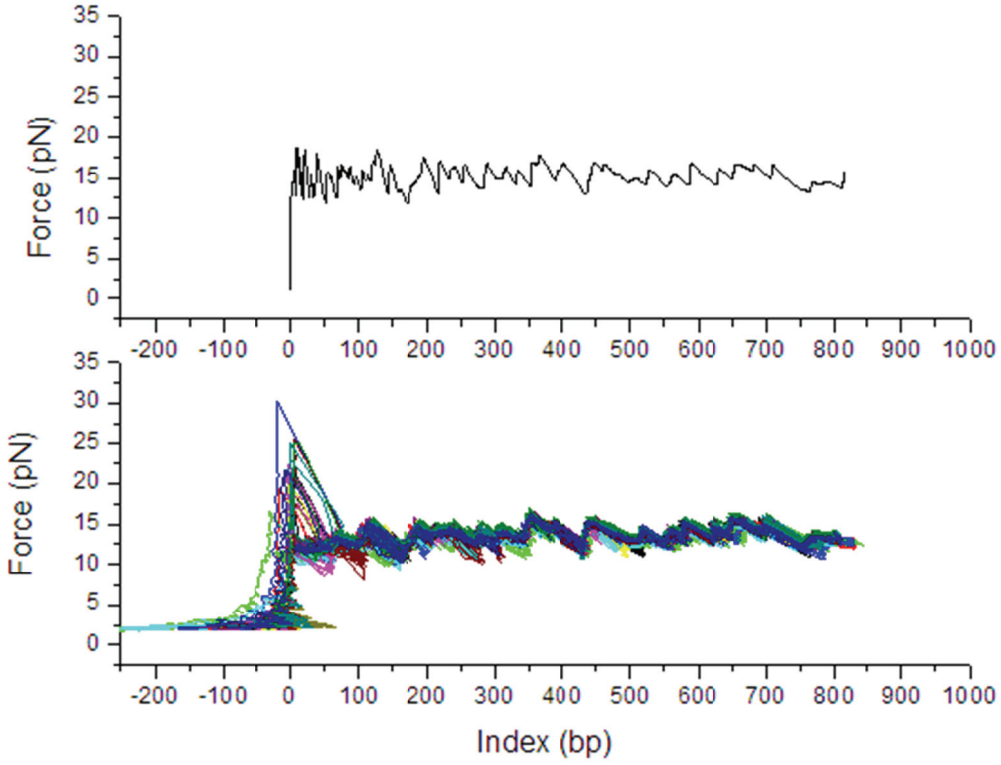

Figure 16) Force vs. basepairs for an unzipping construct made from the plasmid pBR322 in a constant extension unzipping experiment simulation. Top shows simulation, Bottom shows experimental traces.

This is an extremely powerful method for predicting the outcome of real unzipping experiments. This is demonstrated in Figure 16 in which multiple experimental unzipping traces on the bottom panel can be compared with the equilibrium prediction from a constant extension simulation in the top panel. Most features are replicated exceptionally well.

Because values for the force, basepairs and extension are calculated for every position in the simulation, other important relationships can also be predicted, and these may be more interesting for experiments such as the helicase unwinding assay. For example, Figure 17 shows the relationship between unzipped basepairs and extension for the sequence Poly(GC)100 described in the next chapter

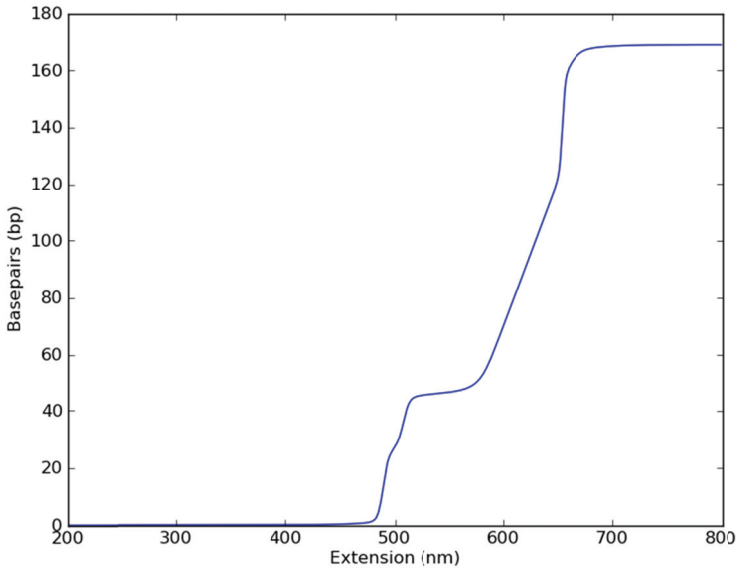

Figure 17) Equilibrium unwound basepairs versus extension for Poly(GC)100

This shows only information about the equilibrium position. If we wish to know more about the probability distribution we can look at an intensity plot showing the probability density of each possible number of bases versus the total trap separation. An example of this is shown in Figure 18. This shows significantly more structure than the equilibrium calculation.

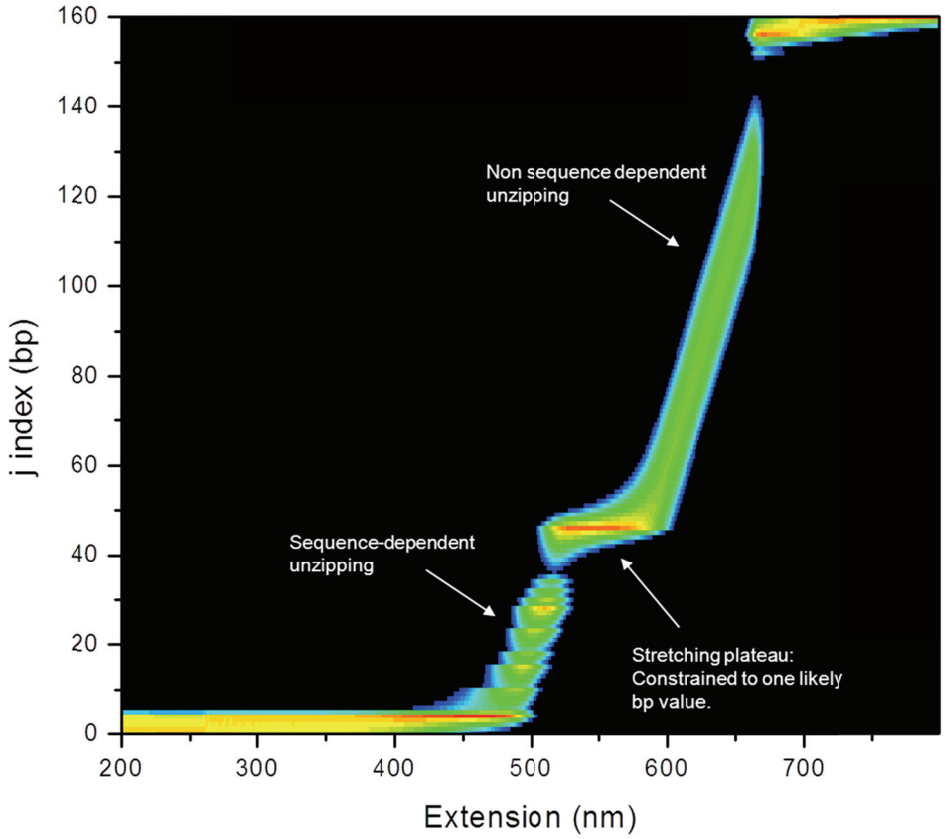

Figure 18) Probability density of basepairs unzipped versus extension. This illustrates highly sequence-dependent regions, a region of no unzipping but stretching, and non-sequence dependent unzipping. Red indicates the highest probability.

Constant Force Experiments

For experiments in which the force is the independent parameter the free energy takes a different form, now including a reduction in the free energy due to the DNA extending into a potential well:

$$\Delta G(n,F) = \sum_{i=0}^{n} \Delta G_{bp}(i) - 2nF.x(F) + 2n\int_0^{x(F)} F(x')dx'$$
$$= \sum_{i=0}^{n} \Delta G_{bp}(i) - 2n\int_0^{F} x(F')dF' \qquad (8)$$

This version of the free energy is appropriate for experiments such as helicase unwinding, in which the force is independently maintained by modulating the trap power via a feedback loop,

Since the force is constant, the integral term will simply increase linearly with the number of basepairs *n*, with the slope determine by the force. This results in more straightforward energy curves for the constant force case, as shown in Figure 19 for Poly(GC)100 at 12pN.

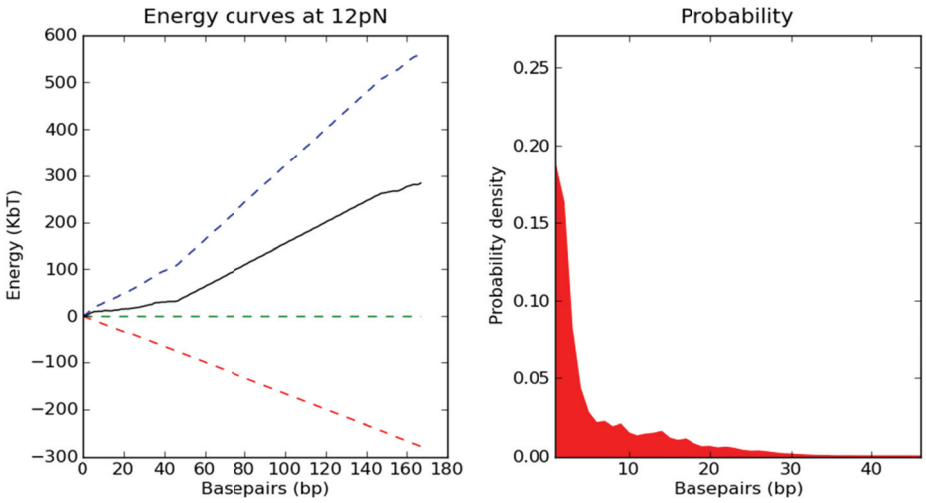

Figure 19) Energy and probability density curves for a constant force experiment at 12pN. Left panel shows ssDNA (red); dsDNA (green); sequence (blue) and total (black) energies. The right panel shows the probability density as a function of unzipped basepairs.

As the force increases the slope of the stretch energy contribution decreases, rotating the total energy closer toward the x-axis as shown in Figure 20:

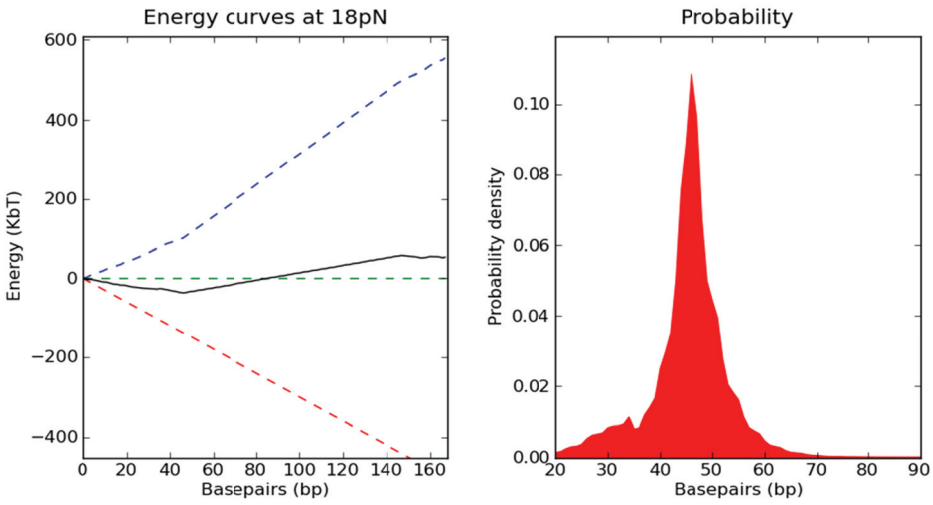

Figure 20) Constant force energy and probability density curves at 18pN

As the force increases the total energy rotates closer and closer to the x axis, until at the critical force the last base will pass though y=0. At this force the construct will spontaneously unzip through to the end. The dsDNA contribution is flat because at a constant force this will always store exactly the same free energy, regardless of the number of unzipped basepairs.

The Effect of Helicase

Both of the types of configuration described (both constant-extension and constant-force) can be modified to account for the addition of helicase unwinding. We will consider first a situation in which the helicase has reached a certain base, *m*. The helicase prevents bases at an index lower than its position from re-annealing, and it also affects the bases ahead of it (M. D. Betterton & Julicher 2003). This is due to the

dissociation potential which allows it to unzip DNA “semi-actively”. For this work I model the potential identically to our previous paper: over a certain range of basepairs the helicase applies a “staircase” potential, with each step equal in height to the energy E0.

Given a certain position of the helicase on the ssDNA we can then calculate the expected position of the fork, by averaging over all allowable open base configurations, but excluding all the bases prevented from re-annealing by the helicase, and with an additional energy term due to the dissociation potential.

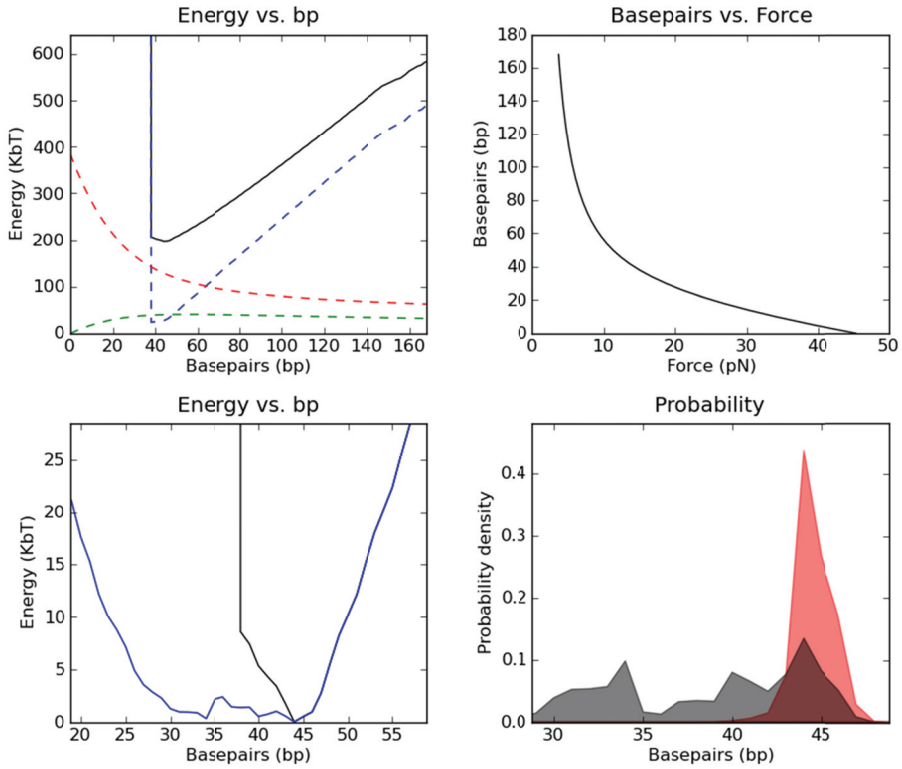

Figure 21) Constant extension with helicase. In bottom left blue is without helicase, black is with helicase. In bottom right black is without helicase, red is with helicase.

Figure 21 shows energy calculations both with and without helicase present for a sequence at a constant extension. In the top left panel the full energy curves with the addition of helicase are shown. In the bottom left panel an enlarged total energy plot is shown around the minimum value, both with the helicase present (black line) and without (blue line). Because helicase prevents the bases from re-annealing, and because the potential acts to decrease the base-pairing energy immediately ahead of the helicase, the potential well around the minimum is far steeper than without the helicase. This results in a significantly tighter probability distribution than with no helicase present, shown in the bottom right panel.

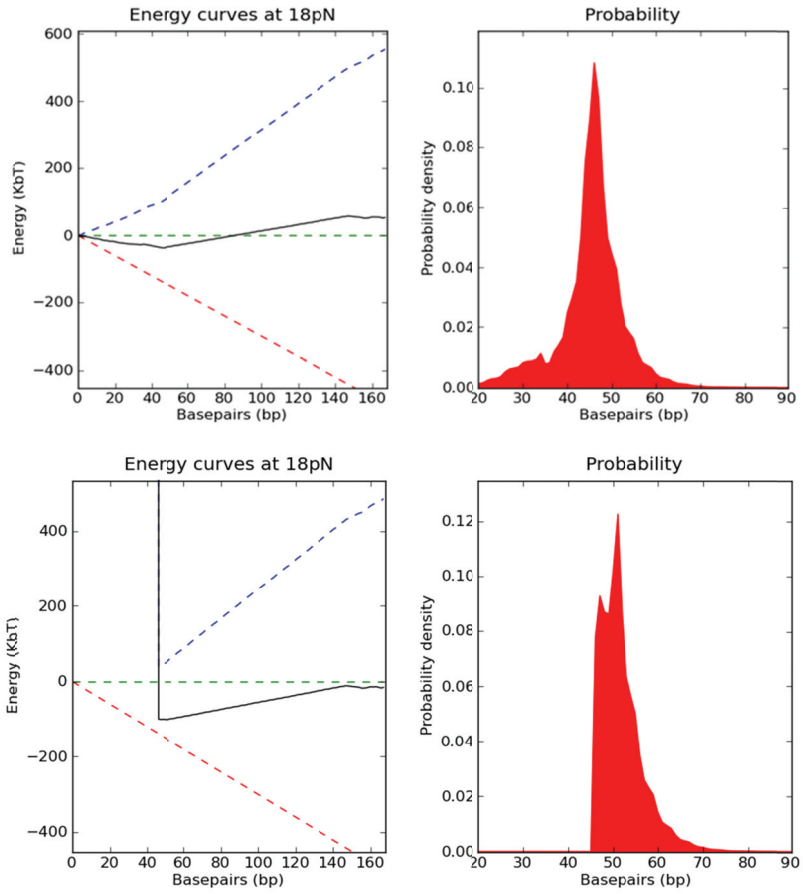

Figure 22) Helicase in constant force assay. Top is with, and bottom without helicase.

For constant force there is also a pronounced effect. This is shown in Figure 22 for a construct held at a force close to the critical unzipping force.

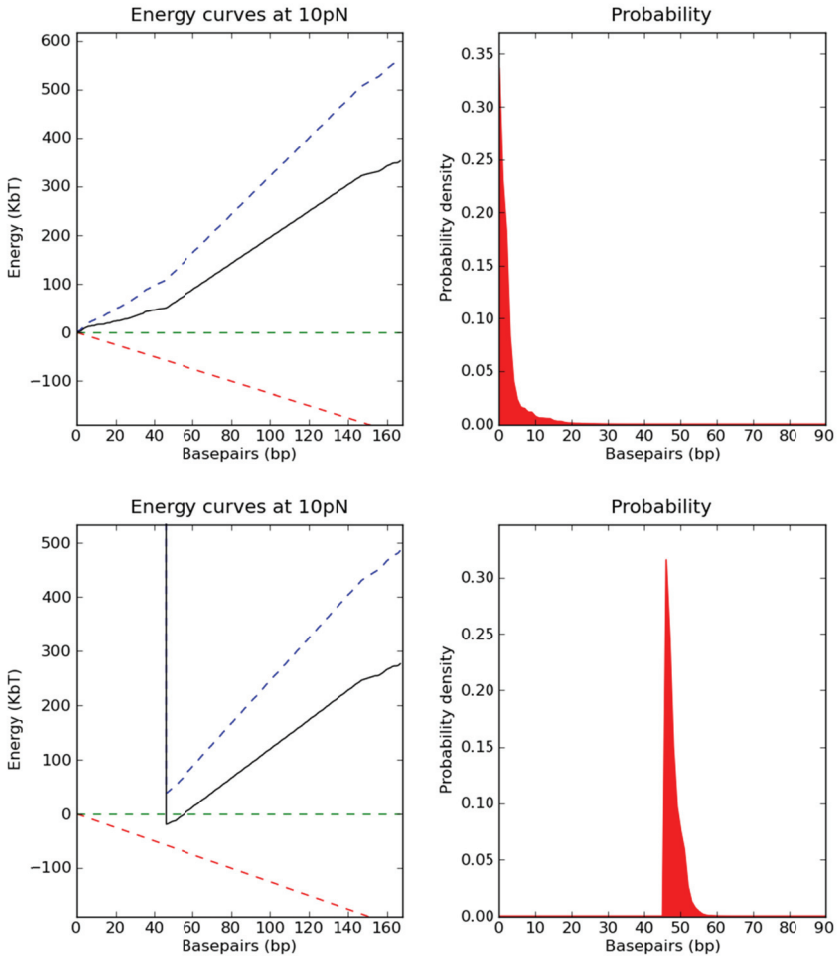

Figure 23) Constant force at 10pN without (top) and with (bottom) helicase present

Adding helicase to a constant force configuration in which the force is well below the critical force has an even more dramatic effect. This is because at a low force the total force can be an extremely steep, monotonically increasing function. Ordinarily this would result in the most likely number of unzipped base being extremely close to zero,

however with the addition of helicase many bases can be prevented from annealing. An example of this for the same construct under a 10pN force is shown in Figure 23.

Magnitude of stepping signal

To determine the magnitude of the step signal we can calculate the response to the helicase moving forward by one physical step. The feedback rate of the system is on the order of 1KHz and the DNA basepairing fluctuations are several orders of magnitude faster (M. D. Betterton & Julicher 2003). The helicase steps at around <100Hz. This means that prior to the step being taken the system is close to being in a true constant force configuration, however as the step is taken it is instantaneously in a constant extension configuration as the system cannot respond immediately to the change in length.

Immediately after a step the length in basepairs of the ssDNA arms will increase, causing a small drop in force as the beads relax to a new equilibrium position, because the system cannot immediately adjust the force to compensate. This means that size of the stepping signal corresponds to this small displacement in the trap of the beads. The magnitude of this step can therefore be calculated by finding both the initial and final positions of the beads in the trap, given a certain set of parameters.

This position can be found by performing an optimization on the equilibrium position of the unzipping fork (using a calculation similar to that described above) and the extension of the ssDNA (this is necessary because the extension determines the force, which determines the equilibrium number of basepairs, which in turn determines the extension). Performing this procedure at every location on a DNA construct gives the

size of the step signal as a function of helicase position. Neglecting stacking energy we see the following dependency:

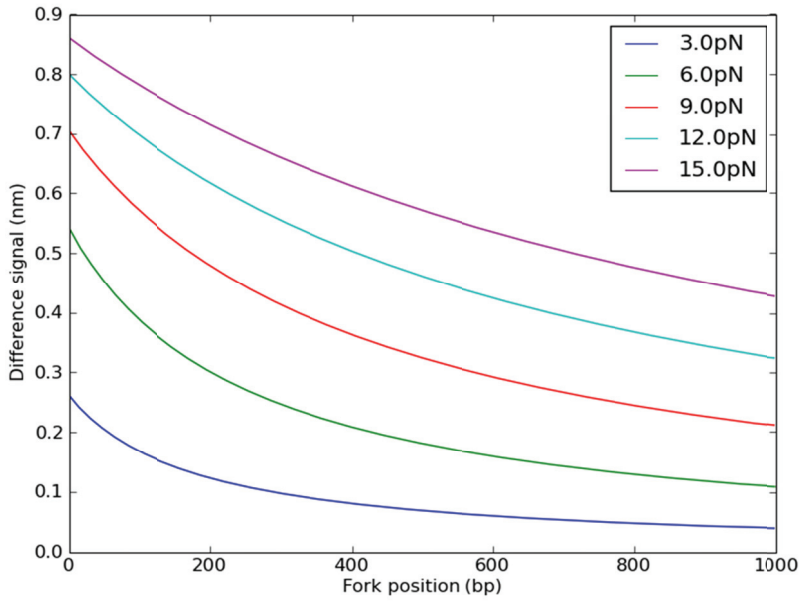

Figure 24) Magnitude of the difference signal for five different forces over the first 1000 basepairs of unzipping, for a homopolymeric GC sequence with a 1bp step-size.

Unsurprisingly this shows that the larger the force, the larger the signal is. This is because the ssDNA is stretched more at higher forces, and the stiffness is greater. Additionally, as the number of bases already unzipped increases, subsequent bases have a progressively smaller effect.

Magnitude of noise

The noise signal occurs due to the Brownian fluctuations of the bead in the trap, modulated by the stiffnesses of the dsDNA and ssDNA, and by hydrodynamic coupling between the beads. Moffitt et. al. perform a detailed calculation to determine the value for the noise signal in a similar, simpler experimental configuration (Moffitt

et al. 2006). We can apply this to the current configuration by modfying their dsDNA stiffness to reflect the combined stiffness of the ssDNA and dsDNA constructs in our experimental setup:

$$k_{total}(F,bp) = \frac{k_{ssDNA}k_{dsDNA}}{k_{ssDNA} + k_{dsDNA}} \tag{9}$$

Where we calculate stiffness by taking the first derivative of the spline interpolations of the force-extension curves. This gives the following expressions for the various noise contributions:

$$\left\langle x_1^2 \right\rangle = 4k_B TB \frac{\gamma_1 \left(k_2 + k_{total}\right)^2 + \gamma_2 k_{total}^2 - 2\frac{\gamma_1\gamma_2}{\Gamma} k_{total}\left(k_2 + k_{total}\right)}{\left(1 - \frac{\gamma_1\gamma_2}{\Gamma}\right)\left(k_1k_2 + k_2k_{total} + k_1k_{total}\right)^2} \tag{10}$$

$$\left\langle x_1 x_2 \right\rangle = \frac{4k_B TB}{\left(1 - \frac{\gamma_1\gamma_2}{\Gamma}\right)\left(k_1k_2 + k_2k_{total} + k_1k_{total}\right)^2} \times \left\{ \begin{array}{l} \gamma_1 k_{total}\left(k_2 + k_{total}\right) + \gamma_2 k_{total}\left(k_1 + k_{total}\right) - \\ \frac{\gamma_1\gamma_2}{\Gamma}\left(k_1k_2 + k_2k_{total} + k_1k_{total} + 2k_{total}^2\right) \end{array} \right\} \tag{11}$$

Where B is the measurement bandwidth, k_1 and k_2 are the stiffnesses of the two traps, γ_1 and γ_2 are the drag coefficients of the two microspheres, and Γ describes the hydrodynamic coupling between the microspheres at distances larger than the bead radius (Meiners & Quake 1999). These are approximately given by:

$$\gamma_{1,2} = 6\pi\eta r_{1,2}$$
$$\Gamma = 4\pi\eta R_{12} \tag{12}$$

Where R_{12} is the total separation of the microspheres and η is the viscosity.

Given these equations for the noise of the individual traps and for the correlated noise, the standard deviation of the noise in the difference signal is given by:

$$\langle x_{diff} \rangle = \sqrt{\langle x_1 \rangle^2 + \langle x_2 \rangle^2 - 2\langle x_1 x_2 \rangle} \tag{13}$$

The magnitude of the noise signal over an appropriate range of forces and basepairs is shown in Figure 25.

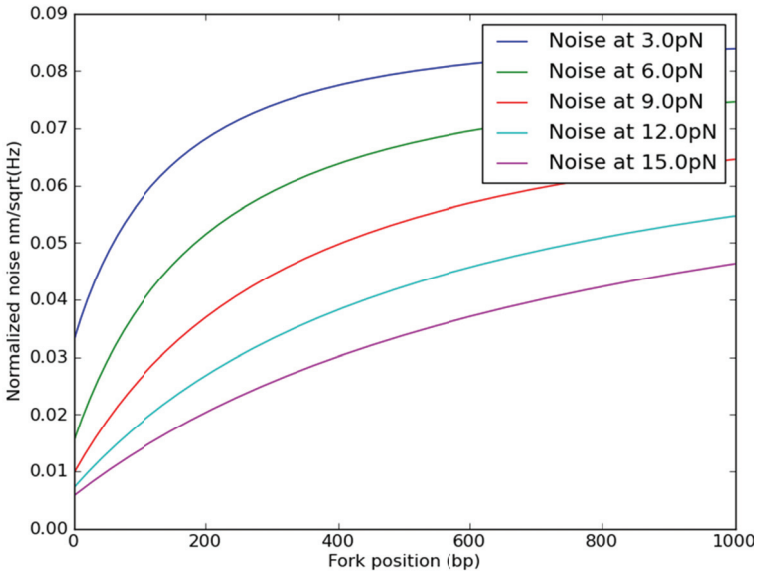

Figure 25) Magnitude of normalized noise signal in the difference co-ordinate for five typical forces, over the first 1000 basepairs

This shows a saturation of noise signal as the number of basepairs increases, and hence the stiffness decreases. This is because the beads get less strongly coupled. The length of DNA required before this is the case is greater for higher forces, because the decrease in stiffness with length is ameliorated by an increase due to the larger force.

Signal to Noise Ratio

This can then be used in combination with the signal simulations described above to calculate the amplitude signal to noise ratio for a specific set of parameters, over a certain measurement bandwidth. This is defined as the ratio of the amplitude of the stepping signal to the amplitude of the noise signal. This shows that the SNR is inversely related to the number of unwound bases. It is also best at high force (Figure 26), with short dsDNA linker segments (Figure 27), with small beads (Figure 28), and large step sizes (Figure 29)

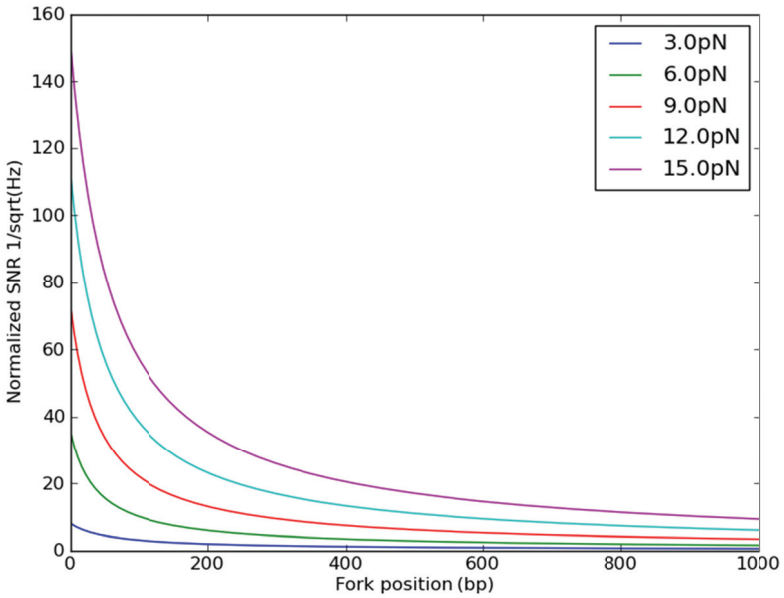

Figure 26) Normalized SNR for five forces over the first 1000 basepairs. Higher forces are better because DNA is stiffer and motion of the beads more coupled.

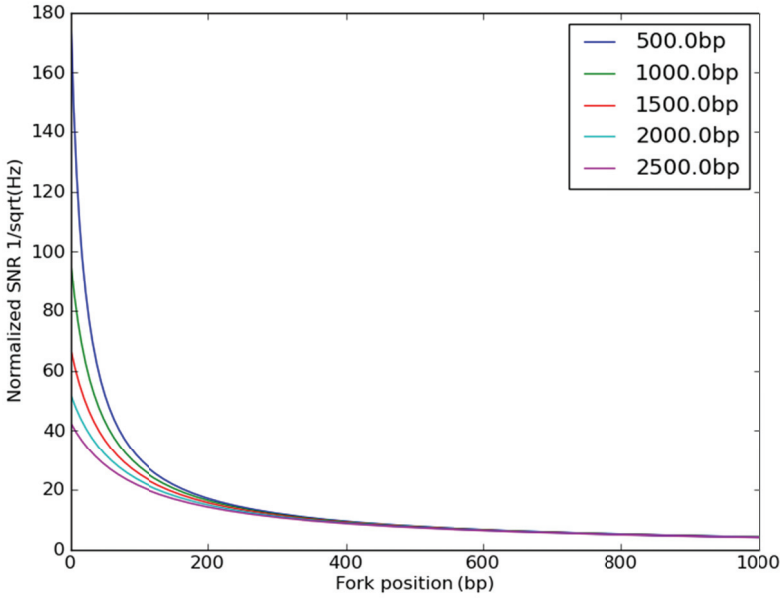

Figure 27) SNR for five different dsDNA linker lengths at F=10pN. Shorter DNA is better because it is stiffer and the beads are more strongly coupled.

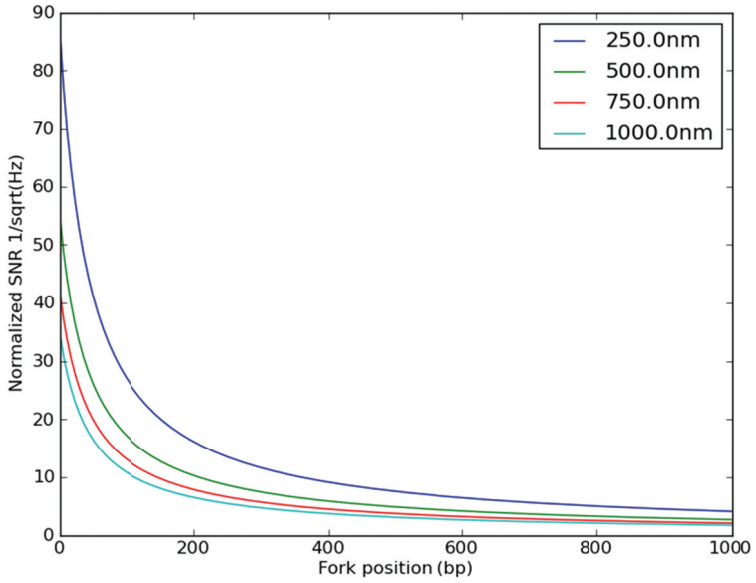

Figure 28) SNR for 4 bead sizes at F=10pN. Smaller beads are better as the drag coefficient is reduced.

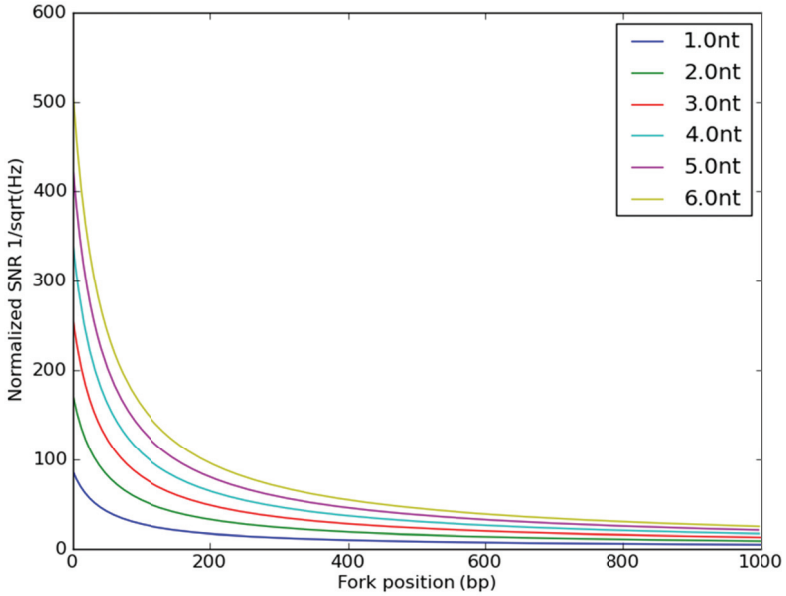

Figure 29) SNR for 6 helicase step sizes at F=10pN. Larger step sizes are better as the signal is larger.

What is an acceptable signal to noise ratio?

In order to determine what range of SNR allows us to reliably extract step sizes I used the information obtained from the simulations above to simulate traces from a helicase stepping experiment.

In general it is advantageous to filter as much noise as possible. Typically this is done in several stages:

1) When collecting data we analog filter at half the acquisition rate. This is to avoid any anti-aliasing from frequencies higher than the collection rate

"wrapping around". We typically use an 8-pole Butterworth filter set at 5kHz, since we collect data at 10kHz.

2) During the data collection the data is averaged and decimated. These rates can be specified by the user.
3) During data analysis we typically will further filter the data using a Gaussian filter in order to further reduce the noise content of the signal.

To simplify this discussion the averaging step will be ignored and it will be assumed that we collect all the data at 10kHz. The effect of the 2^{nd} and 3^{rd} averaging steps can be combined by performing a single low pass filter.

The procedure to use the calculated SNRs to simulate unzipping traces is as follows:

1) Find the total amplitude noise signal integrated over the measurement bandwidth.
2) Generate Gaussian distributed noise for every point in the simulated signal
3) Generate a simulated stepping trace given the average helicase translocation rate.
4) Add the signal and noise signals together
5) Apply digital low-pass filtering

For example: under typical conditions and 12pN of force we get a SNR of about 50 $nm/Hz^{0.5}$ for 1bp steps, for the first 200bp. Steps are simulated by choosing rates from an exponential distribution with an average rate of 50Hz, with 10,000 points generated for each second of simulated data, mimicking the experimental data collection rate. This signal is then added to a simulated noise signal generated randomly from a

Gaussian distribution with a standard deviation of 0.63, corresponding to the SNR at the Nyquist frequency. This reproduces the noisy stepping signal after being anti-alias filtered and sampled by the data acquisition hardware. Next the signal is resampled (to increase processing speed) and low-pass filtered to a frequency of twice the average stepping rate. This procedure, with the above parameters is illustrated in Figure 30.

When simulating the helicase steps it is assumed that the stepping rate follows an exponential distribution. This is valid if there is a single rate limiting step in the biochemical cycle of the helicase. Otherwise, it will follow a Gamma distribution (K. C. Neuman et al. 2005). To simulate this data rates are generated by using

$$k = \frac{-\ln(x)}{k_0} \tag{14}$$

Where x is an evenly distributed random number between 0 and 1. An example of a stepping signal generated using this procedure is in Figure 30a. This demonstrates the distribution of plateau sizes due to the exponential distribution of rates. Some steps remain fairly obvious after processing, but there are also sections in which multiple steps occur in very close in time, and are hard to distinguish in the processed data.

There are many methods to retrieve steps from noisy data. A recent paper (B. C. Carter et al. 2008) comparing many of these methods found that the iterative χ^2 step-fitting procedure proposed by Kerssemakkers et. al. offered the best performance (Kerssemakers et al. 2006).

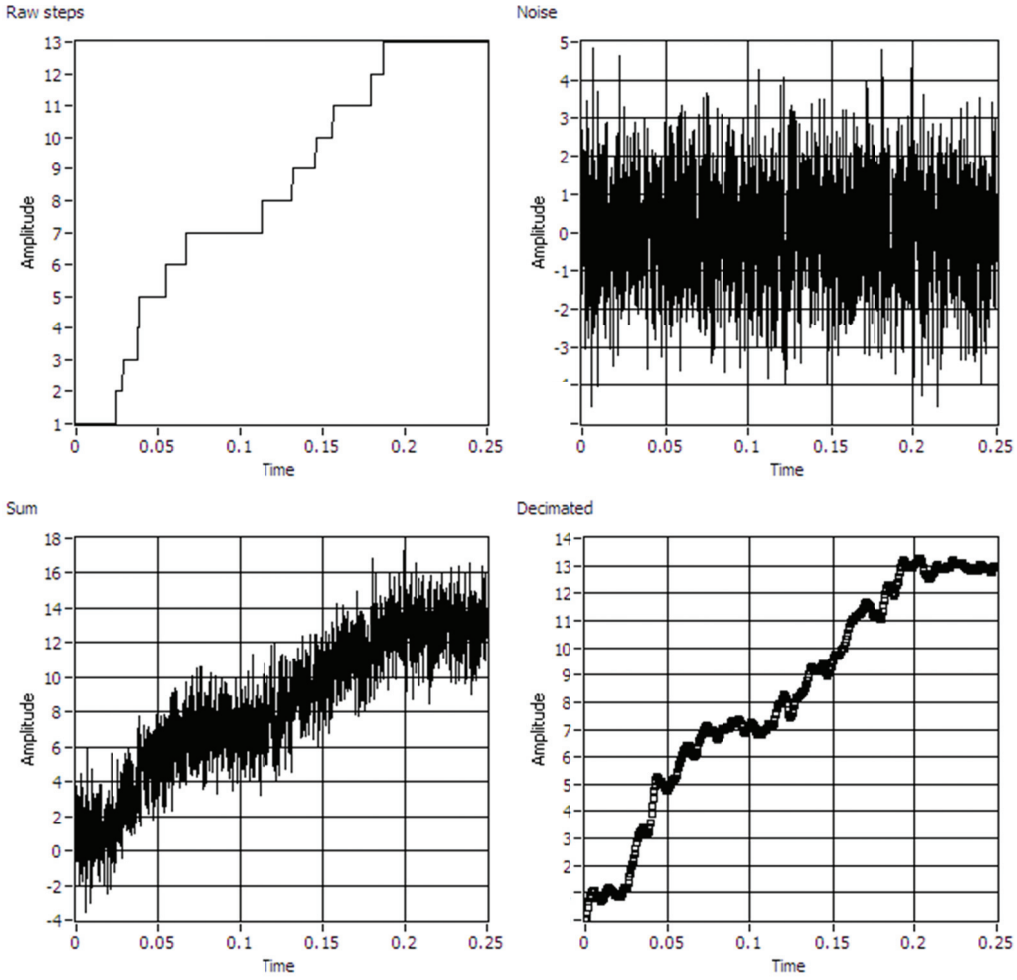

Figure 30) Illustration of construction of simulated unzipping traces. a) Simulated unzipping signal from exponential distribution with average rate of 50Hz; b) Gaussian noise with σ=0.63; c) Sum of signal and noise; d) Sum filtered to 100Hz

The Kerssemakers step-detection algorithm works by fitting a single step to the data in the most "significant" location. The significance, a metric specific to this algorithm, is determined by multiplying the square root of the number of points in the step by the height of the step. Then, steps are fit to all the resulting plateaus, and the most significant of these is retained.

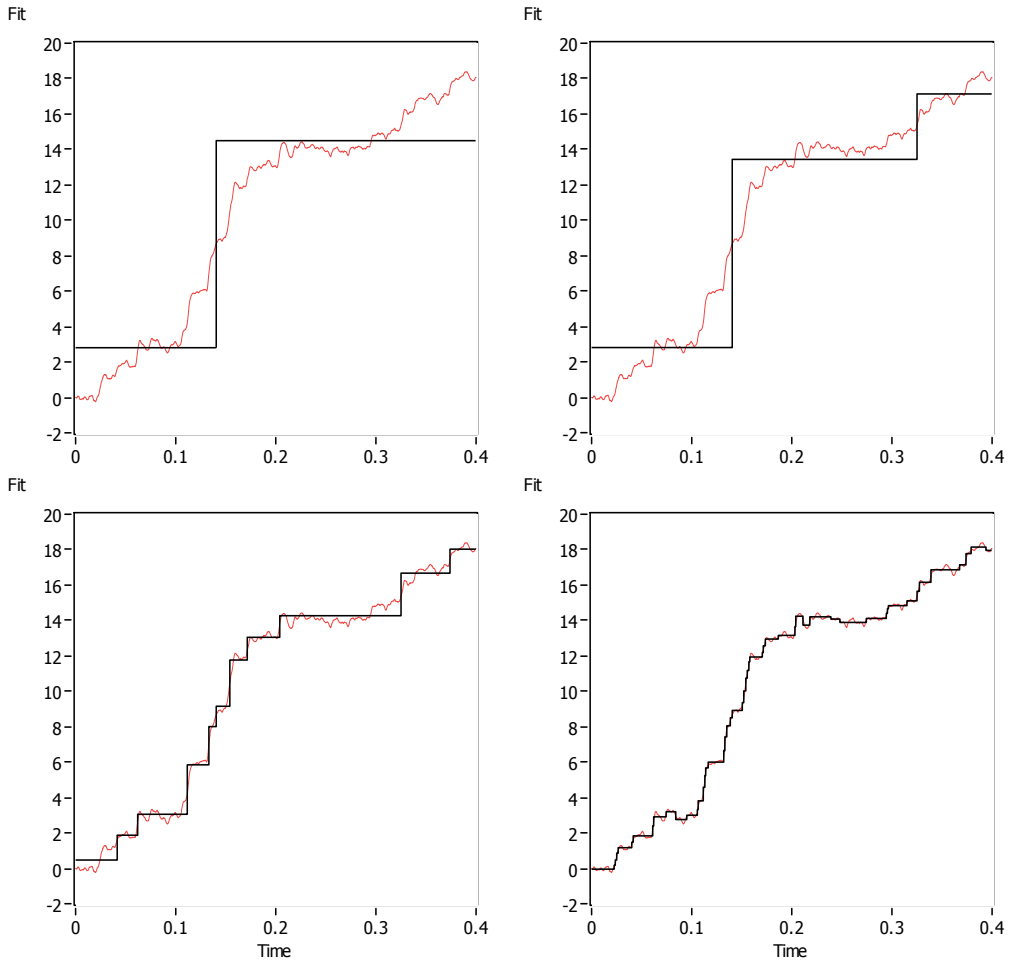

Figure 31) Example fits for 1, 2, 10 and 50 iterations of the Kerssemaker step fitting algorithm. After 50 steps the noise is fit rather than true steps.

Eventually as the steps get smaller and smaller they will start fitting the noise instead of steps. In order to determine when this is the case, at each fitting iteration a "counter" fit is performed in which alternative steps are fit between the plateaus of the existing steps, and all the existing steps are then removed. If it is purely noise remaining, both of these fits should work well. If instead the correct number of steps are being used to fit the data, an alternative fit should fit significantly less successful.

By comparing the goodness of fit between the regular and counter fits one can find a peak at the correct number of steps (Kerssemakers et al. 2006).

Another, simpler step detection method is the pairwise distribution. In this method every point of data is subtracted from every other point, and these differences are then binned on a histogram. If there are recurring steps the histogram will show a peak at multiples of the step-size (B. C. Carter et al. 2008).

By experimenting with the simulated data, for conditions which mimic reasonable experimental conditions, the Kerssemakers algorithm could identify steps at slightly lower SNR's than a simple pairwise distribution, as shown in Figure 32.

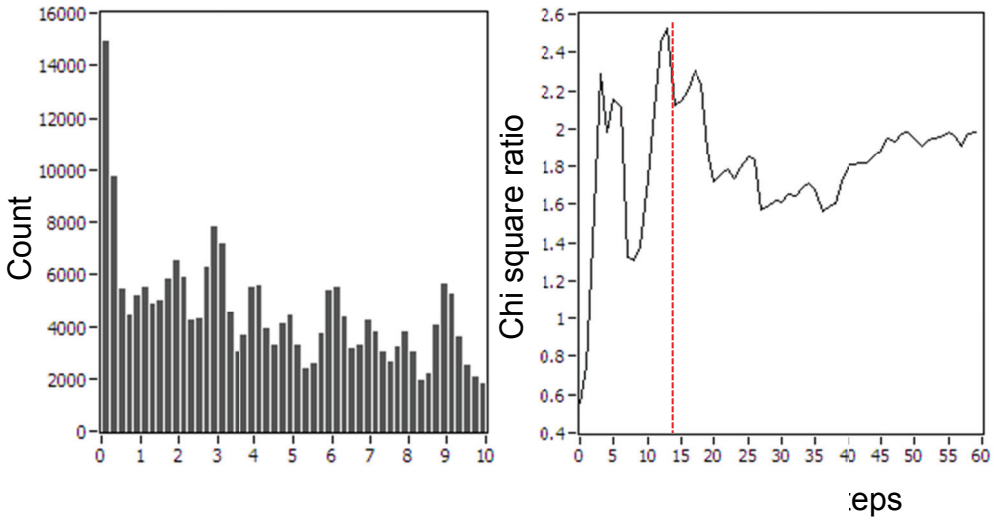

Figure 32) Pairwise distribution (left) and Kerssemakers fits (right) at SNR=50Hz$^{-0.5}$. The red line indicates the true number of steps, matched well by the peak.

However, it did not appear to be substantially more powerful, and it is occasionally difficult to interpret, especially for data with few steps. By experimenting with SNR's it is evident that the point at which it is hard to discern steps using either of the two identification algorithms described occurs at about an SNR of $50Hz^{-0.5}$, for our experimental parameters, with an average helicase rate of 50Hz. This is demonstrated in Figure 33.

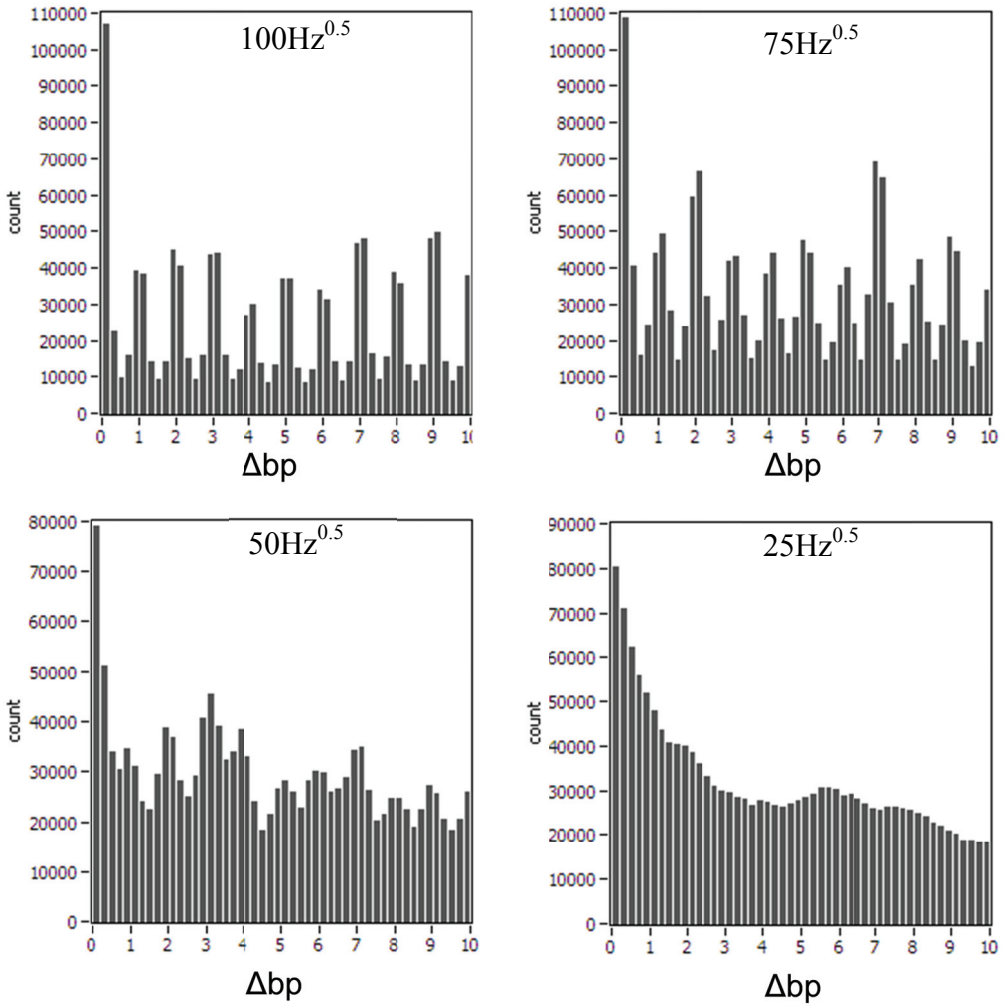

Figure 33) Example pairwise distributions for simulated traces at various SNRs for helicase moving at 50Hz and signal filtered to 100Hz.

Comparing this SNR value to the calculated SNR values from simulation illustrates that under optimal experimental conditions -- 1000bp dsDNA tether, 250nm radius beads – the SNR is almost precisely at the threshold for step detection. Because this is an idealized calculation of noise there will be many other sources present during a real experiment. This means in order to maximize the ability to see 1bp steps the helicase likely needs to be slowed to rates slower than 50Hz, as well as optimizing all of the experimental parameters discussed above.

Summary

ssDNA and dsDNA have a well-characterized, non-linear elastic response. Using analytical expressions for these relationships, in combination with knowledge of the base-pairing energy and sequence, equilibrium simulations of constant-extension and constant-force experiments can be performed. These are very successful in predicting experimental data. Helicase can be added to the simulations and the magnitude of the signal generated by a single unwinding event can be calculated. The noise can be calculated using the instantaneous stiffness of the system and hydrodynamic coupling of the beads in solution. This allows a calculation of the SNR under specific experimental conditions. By analyzing simulated stepping data the threshold SNR for observing steps can be determined. The simulations indicate that under optimal experimental conditions, single base pair stepping is detectable at stepping rates of less than ~50Hz.

Chapter 4:
Measuring the Step-Size of T7 DNA Helicase

Introduction

To probe the step size of T7 DNA helicase I used a dual trap configuration like that shown in Figure 9. Obtaining this configuration with the DNA stretched between the traps is significantly more challenging than in the single trap configuration, particularly when the trap is built around a microscope. In order to accomplish this a novel method for forming double tethers was implemented, using a method of temporary attachment to the sample chamber surface. Using this method discrete steps of multiple sizes were observed as T7gp4 unwound a GC rich construct, the smallest of which was 1bp.

Forming tethers in single trap configuration

In the single trap configuration it is straightforward to prepare a DNA chamber. A typical protocol used is:

1. Incubate sample chamber with anti-digoxegenin diluted in PBS for 12 minutes
2. Wash chamber twice times with 4mg/ml casein solution
3. Incubate chamber with DNA for 12 minutes
4. Wash chamber twice with 4mg/ml casein solution
5. Incubate chamber with streptavidin coated beads for 12 minutes
6. Wash chamber twice with 4mg/ml casein
7. Flow in helicase solution

The anti-digoxegenin binds non-specifically to the surface, providing an attachment point for the digoxigenin tagged DNA. Casein is a blocking agent, which prevents DNA and beads from sticking to the surface by coating it with a hydrophobic layer.

The streptavidin coated beads attach to the second tag on the DNA which is a biotin modified base.

When this chamber is observed through the microscope, successfully formed tethers will be indicated by the presence of beads loosely coupled to the surface. The beads show Brownian motion due to the flexibility of the DNA tether, and the length of the tether may be estimated by observing the degree of motion (P. C. Nelson et al. 2006). To perform experiments one simply centers the trap over the bead and begins the experimental process.

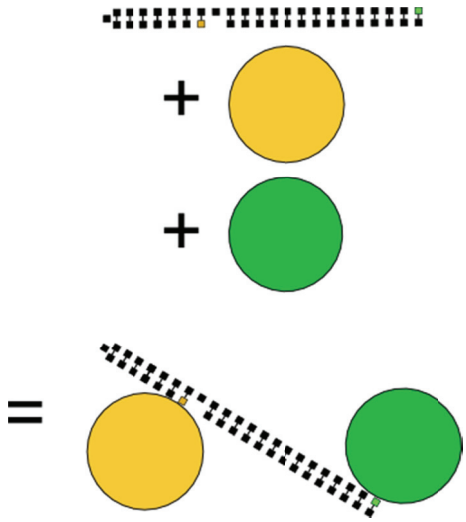

Figure 34) Double-tether unzipping construct. One of each coated bead, anti-dig (green) and streptavidin (yellow) are combined with an DNA unzipping construct to form a double tether.

Forming tethers in a dual trap

In a dual trap configuration it is more complicated to form tethers. This is because by definition the construct is not attached to the surface of sample chamber once the

experiment is being performed. However, the DNA construct must still somehow be loaded into the pair of traps. The ultimate product will be formed from one of each type of bead, linked with a DNA construct, this is shown schematically in Figure 34.

One common solution is to use a flow cell device (Brewer & Bianco 2008). In these devices there are typically three distinct regions. One region contains one of the two varieties of beads, a second region contains the other variety of beads with DNA attached at one end, and the third region contains the experimental buffer. One of the beads can be attached to the DNA by incubating them together before they are flowed into the chamber (Landry et al. 2009).

Tethers are then formed within such a device by manually combining the components. First, one of the beads is trapped in one of the two traps. Then the chamber is moved so that the other type of bead can be trapped within the second trap. Next the two trapped beads are moved into the experimental buffer region. At this point the beads can be brought together so that the DNA strand attached to one can attach to the other via the second linkage. Successful linkage can be observed by pulling the beads apart and observing whether the motion is coupled.

There are several difficulties which may be encountered when using a flow cell configuration. First, there must be a method of delivering buffer and beads to each of the channels. This requires a more complex sample chamber design, such as adding holes through the coverslip or slide, and additional components outside of the chamber itself. These additional components are typically bulky pieces of equipment such as capillary tubes, pumps, regulators and vacuum connectors. Space will often be at a premium in experimental optical trapping setups, especially those built around a

microscope body. This is because the condenser lens covers a very wide area, making it difficult to add to the size of the sample chamber.

Another difficulty with a flow cell is that material must be cycled through the device, requiring much higher volumes of sample buffer than single trap experiments. While this may not be a problem for experiments which require a relatively simple and inexpensive buffer, it can potentially be extremely expensive or impossible for experiments in which the buffer contains purified proteins.

A final difficulty with flow cell designs is the process of tether assembly. Because this process requires bringing the two beads close together it is possible for the beads to be pushed out of the traps, or both pulled into the same trap. This becomes more likely as the DNA tether length is decreased as this necessitates a close approach between the traps.

SLUDTS: Surface-Linked, Unzippable Double Tethers

To avoid the complications associated with flow cells I developed a novel technique for forming double tethers. This method attaches double tethers to the surface of the sample chamber using a temporary connection, allowing the experimenter to "peel off" ready-made double tethers from the surface.

For tethers to be able to be removed from the surface, the connection to the surface must break when a force is exerted by the optical traps. If this were not the case it would be impossible to remove the double tethers from the surface. Typical bonding interactions -- such as antibody interactions -- are unsuitable for this because they are too strong, breaking at over 70pN of force (Kienberger et al. 2005).

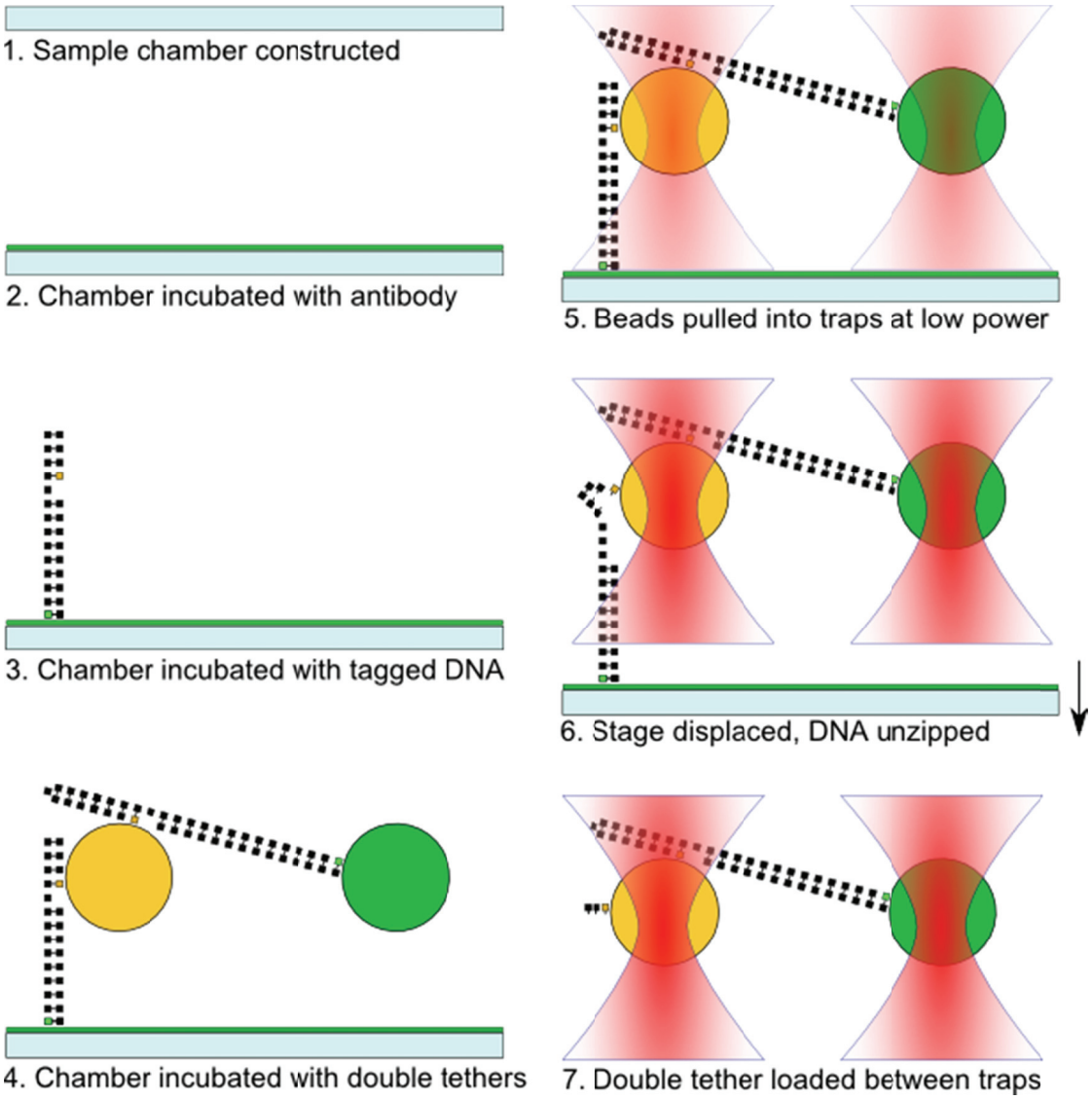

Figure 35) Assembly and removal of SLUDT constructs

Instead of these very strong bonds, a short (111bp) strand of unzippable DNA is used as a temporary connector between the double tether and the surface. By providing enough force to unzip this short construct, the double tether can be unzipped, thereby detaching it from the surface.

The process of removing an unzippable double tether from the surface is shown in Figure 35. They can be identified under the microscope by the presence of the two tethered beads, showing tethered Brownian motion. Upon finding a tether the two beams are first lowered in power, so that the maximum possible applied force is less than the unzipping force. Then the beams are brought to the position of the beads. By pulling the construct sideways, the two beads tend to separate along the X-Y plane because of the geometry of the surface connection. Once each of the beads is in a separate trap the power is increased and the surface displaced in the Z direction. This unzips the short tether connecting the beads to the surface, leaving a dual tether stretched between the two traps.

A typical protocol for forming a chamber is as follows:

1. Mix anti-digoxegenin beads, streptavidin beads and DNA. Incubate for 1 hour with mild agitation.
2. Incubate sample chamber with anti-digoxegenin diluted in PBS for 12 minutes
3. Wash chamber twice times with 4mg/ml casein
4. Incubate chamber with short unzipping DNA for 12 minutes
5. Wash chamber twice with 4mg/ml casein
6. Incubate chamber with pre-incubated mix from step 1 for 40 minutes.
7. Wash chamber twice with 4mg/ml casein
8. Flow in helicase solution

Helicase Step-Size Assay

To probe the physical step-size of T7gp4 a construct containing a long run of purely GC bonds was used, called Poly(GC)100. This reduces the sequence dependent

structure, allowing the steps of the protein to be probed. The construction of this template is illustrated in Figure 36:

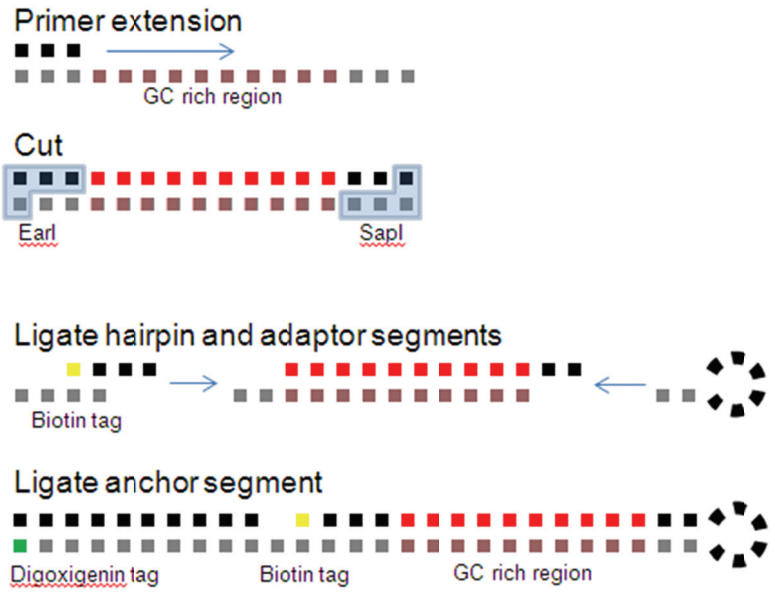

Figure 36) Outline of construction of PolyGC(100)

The construct was prepared by first performing primer extension on a synthetic 145nt template (Integrated DNA Technologies, Coralville, IA). The sequence of this template was rich in GC content, with a higher AT content toward the 3' end:

5'GATCCTAAGAAGAGCCCGCCCGCCCGCCCGCCGCCCGCCCGCCGCCCGC
CCGCCCGCCCGCCCGCCGCCCGCCCGCCCGCCGCCCGCCCGCCGCCCGCCC
GCCCGCCCGCCCGCCTTATTTAAATGCTTGAAGAGTGACCGACG-3'

After performing primer extension the resulting dsDNA was cut using two restriction enzymes, EarI and SapI (New England BioLabs, Beverly, MA) leaving overhangs

which could be ligated to additional pieces of DNA. One end was ligated to a hairpin construct with 5bp in the loop. The other end was ligated to an adaptor segment which contained a base modified with biotin. This intermediate construct was then gel-purified before being ligated to the standard anchor segment described previously (Koch et al. 2002). A final gel purification was then performed so that only the complete construct remained.

The final experimental buffer conditions were 10nM EDTA, 1mM dTTP, 50nM MgCl, 200nM gp4A' monomer, with 0.4mg/ml casein, 0.02% Tween 20, 50 mM NaCl, and 20 mM Tris-HCl (pH 7.5). Although the exact rate of helicase unwinding is dependent on force, at typical forces between 14-18pN this resulted in an unwinding rate of around 20-50Hz.

An example of a helicase unwinding experiment is shown in Figure 37. After a double tether was loaded into the dual trap, approximately the first 50bp are unzipped manually. The force is then maintained at a value between 14-18pN. The helicase could then bind to the exposed single-stranded region. If this occurred it would first translocate along this region and then upon encountering the dsDNA fork would start to unwind the DNA. This would result in a decreasing force on the beads, due to an increasing number of ssDNA bases in the arms. To compensate for this drop in the force the movable trap would be translated away from the fixed trap, increasing the tension in the DNA until the force reaches the set-point. This is maintained until either the DNA is completely unzipped, or it breaks.

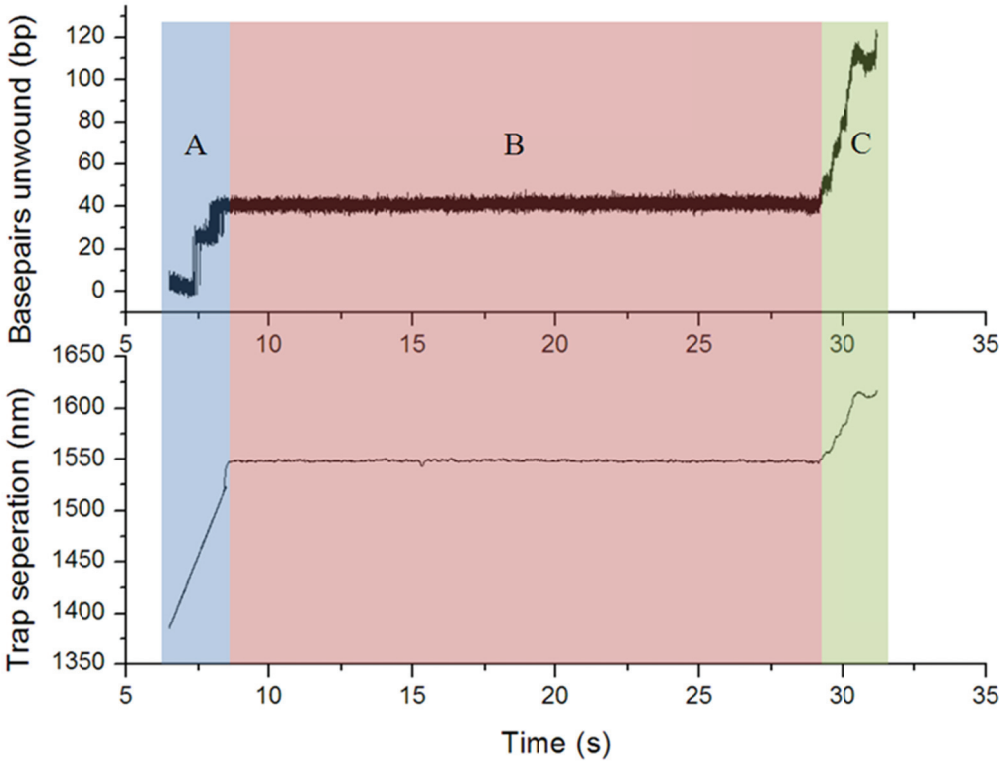

Figure 37) Example of a helicase unwinding experiment. A) Approximately 45 basepairs are manually unzipped using a velocity clamp; B) The force is maintained using a force-clamp, helicase in solution are now able to bind to the exposed ssDNA loading region; C) Helicase unwinds the construct while the force is maintained

From the measurements of the force and trap position the number of basepairs unwound as a function of time can be monitored, and if the signal to noise ratio is high enough, individual physical steps can be observed.

To slow down the rate of helicase translocation, thus increasing the signal to noise ratio of the measurement, we used a reduced magnesium concentration when compared to physiological conditions. A reduction in magnesium concentration from

7mM to 50nM causes the helicase velocity to slow from around 250bp/s to less than 40bp/s when a force of 18pN is maintained.

This method of slowing the helicase was chosen since it has been shown that helicase can bind to DNA without difficulty when magnesium is not present. This indicates that if the metallic cofactor is not present the helicase can be in a stably bound but non-translocating state (Kristen Moore Picha & Smita S. Patel 1998). By increasing the likelihood of entering the paused step and then examining the size of the following step we can probe the step size.

Experimental Results

A representative trace is shown in Figure 38.

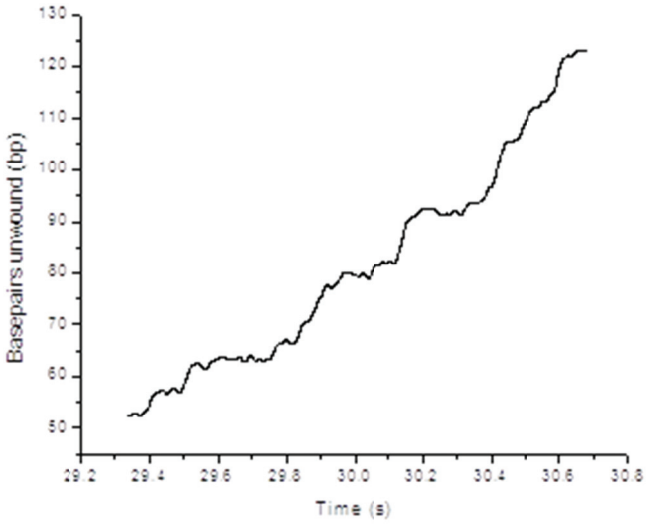

Figure 38) Example of helicase unwinding trace filtered with 8 pole Bessel filter at 75Hz

This shows the calculated number of unwound basepairs as a function of time. The region over which this data is taken is nearly uniform in sequence, but the data clearly shows plateaus. Traces such as this can be analyzed by visual inspection, and using the step identification techniques described above.

A variety of steps ranging in size from single basepairs upward were observed over multiple traces, depending on the level of noise. An example of a trace exhibiting single basepair steps is shown in Figure 39. The black line in the figure shows the raw data digitally low-pass filtered to 40Hz with a Bessel filter. The red overlay shows a Kerssemaker fit to this data. Figure 40 shows the pairwise fit to this data, clearly showing 1bp steps.

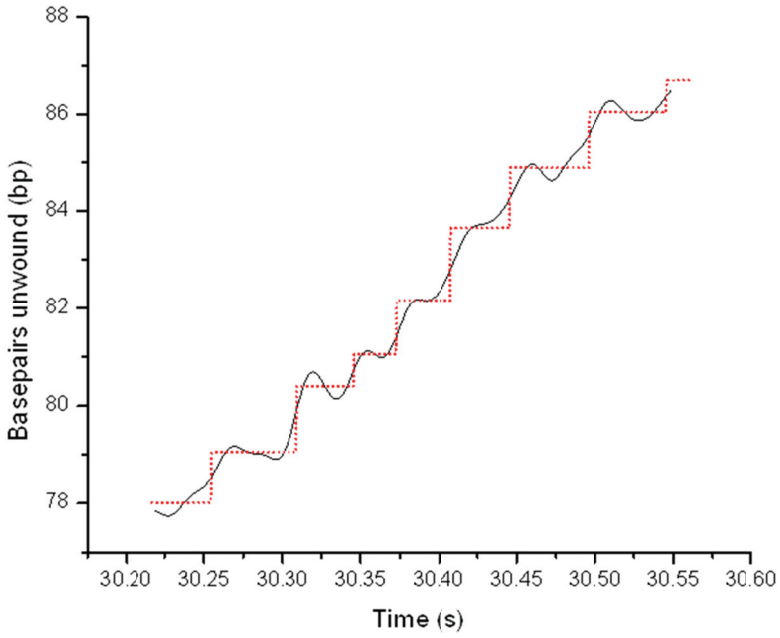

Figure 39) Black line is raw data filtered to 40Hz, red is a fit with the Kerssemaker algorithm.

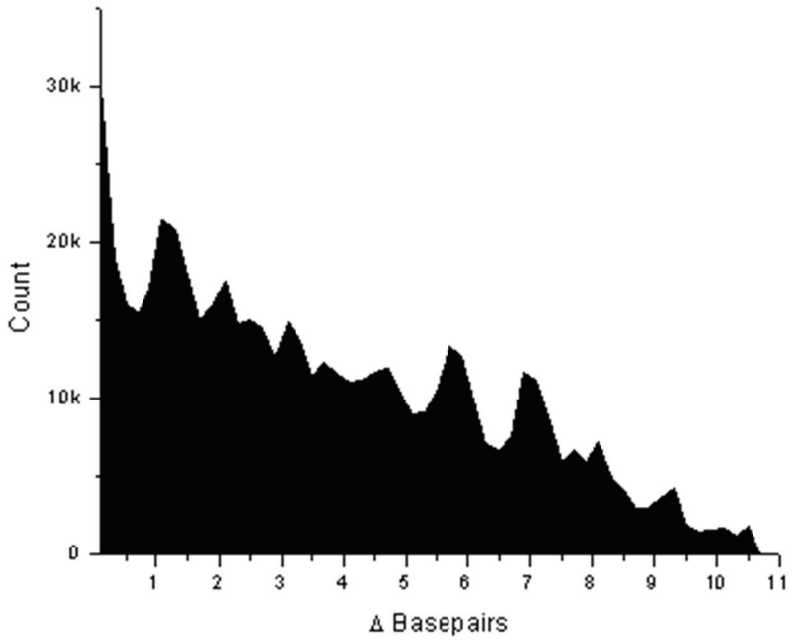

Figure 40) Pairwise fit to the data shown in Figure 39.

Other step sizes were also observed. Figure 41 shows a trace exhibiting a variety of small step-sizes

Interpretation of results

The smallest step observed was 1bp. Many other sizes of step up to at least 10bp were also observed. This observation is consistent with a model in which the smallest physical unit is 1bp, but this may occur in bursts of any multiple of this smallest unit. Because the rates of the individual steps are likely to be distributed in an exponential, or close to exponential manner as discussed earlier, this means that on many occasions the steps will be taken too quickly to be resolved.

.

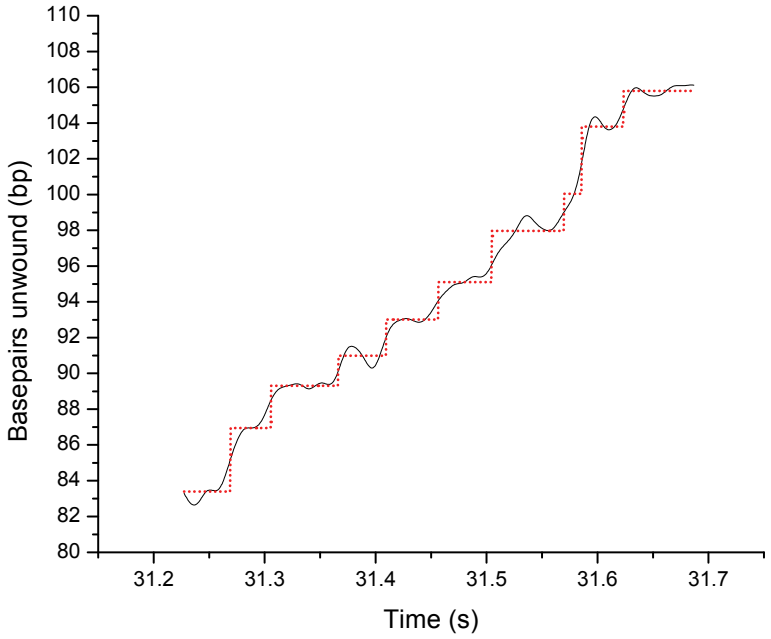

Figure 41) Example trace showing multiple step sizes.

The observation of 1bp minimum steps supports a model for helicase translocation in which the helicase advances due to the action of one subunit at a time. Crystal structures indicate that adjacent monomers are in progressively different states (Singleton et al. 2000; Toth et al. 2003). This makes it highly likely that the mechanism is sequential.

One possible configuration is that each subunit is active in turn. This is illustrated schematically in Figure 42a. One subunit is initially bound to the ssDNA strand in an unrotated position, with a dTTP molecule in the hydrolysis site (state A). Upon hydrolysis the binding loop is rotated downward, pulling the ssDNA through the central channel by 1nt. The conformational change affects the next subunit, altering the next site so that it is more likely to bind to DNA and hydrolyze a dTTP molecule.

In this manner only one subunit is active at a time, and each moves the ring along the DNA by a single physical step, before handing it off to the next subunit.

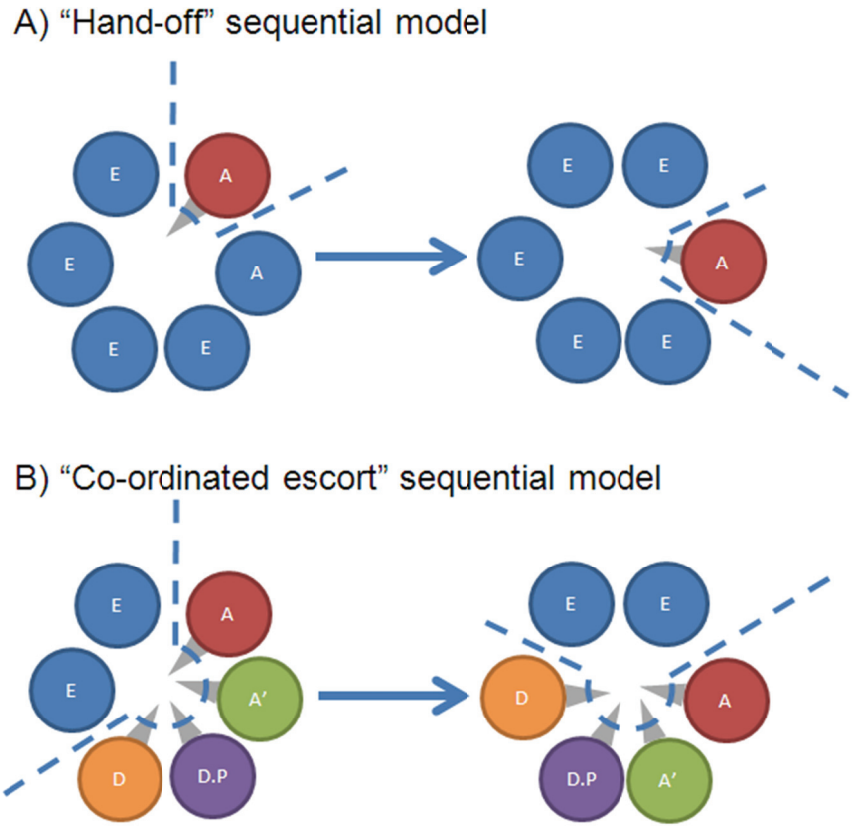

Figure 42) Two sequential models for T7gp4 translocation

Another possible configuration is a coordinated escort model, as proposed for Rho and E1. In this model many subunits are bound to the DNA at once, however each is in a different physical state, and these states progress around the subunits in the ring. One or more of these states do not strongly bind DNA. Each subunit will continuously cycle through each of the possible states, while maintaining its position in the cycle

relative to its neighbors. This results in translocation of the helicase as the DNA is escorted through the ring by the binding of the subunits.

The latter model is supported by the crystal structure of other hexameric helicases crystallized with DNA and RNA (Enemark & Joshua-Tor 2006; Thomsen & Berger 2009), and also by other experiments within our lab which indicate that more than one monomer is bound to the DNA at once. In combination with these results, the observation of single base-pair steps gives strong support to the proposed sequential translocation mechanism, particularly the co-ordinated escort model.

Summary

Double-tether constructs are more difficult to load into optical traps than a single trap configuration. Typically flow cells are used, but the space requirements for this is not always compatible with an existing microscope design. A novel method of loading double-tethers was implemented, by using a temporary connector to the surface consisting of a very short unzipping construct. Using this method of loading the double tethers, single base-pair steps were observed as T7gp4 helicase unwound a GC rich DNA construct. This supports a sequential mechanism with a 1bp physical step size.

Chapter 5:
Future Work

Introduction

There are a variety of further experiments which follow naturally from the work conducted so far. These include experiments which probe and verify the sequence dependence of the helicase stepping mechanism, experiments to probe the loading mechanism of the helicase, and eventually experiments on eukaryotic replicative helicase.

Verification of sequence dependent structure not playing a role

One important control is to verify that the DNA sequence is not artificially imposing steps onto the measurement. Although it is unlikely that this is the case given the size of the steps observed compared to the sequence, it could be explored further through the use of varying sequences. One difficulty with this is that the sequence needs to be purely G and C based in order to maintain a high force and reduce sequence variation, however these sequences tend to form G quadruplex structures (Bryan & Baumann 2010).

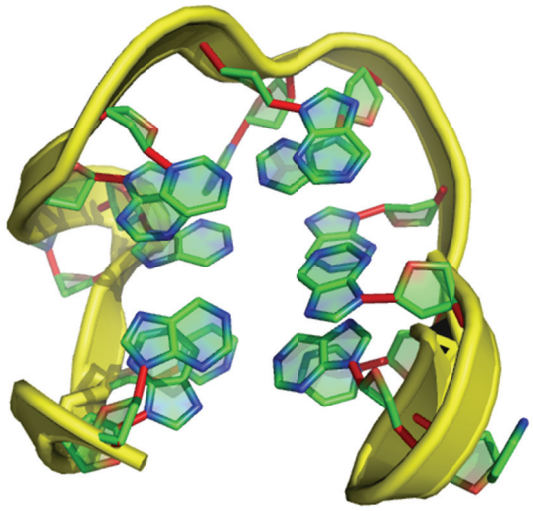

Figure 43) Structure of G quadruplex

These structures are caused by cross-binding of guanine residues, in an alternative fashion to the usual hydrogen bonds between complementary strands. This cross binding is extremely tight, containing many more hydrogen bonds than standard base-pairing and able to withstand forces of many tens of pico-Newtons (Lynch et al. 2009). If these form on an unzipping template it will drastically alter the outcome of any unwinding experiments.

This makes it difficult to create constructs with long repeats of G's. As a general guideline sequences of the form $d(G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+})$ are susceptible to G-quadruplex structure formation (Keniry 2000).

PolyAT

One variation in sequence which would be particularly interesting to try would be a poly(AT) construct. This would be similar to the poly(GC) construct but instead of long runs of guanine and cytosine, it would use the weaker adenine and thymine bases. This would be an interesting experiment as there are predictions that T7gp4 has a different unwinding rate on AT vs GC bases (Donmez & Smita S Patel 2008). It may also reveal information about the method of strand separation due to the weaker base pairing energy of the AT bases.

A major challenge with this construct would be the extremely low unwinding force. A construct containing purely AT bases would mechanically unzip at less than 10pN and all experiments with helicase would thus need to be performed at forces even less than this. As shown by the simulations this causes a significant increase in noise, since the DNA stiffness is so low at these forces.

Loading mechanism

Another area which is related to the helicase step-size measurement is the loading mechanism. T7gp4 is known to form rings before loading onto the ssDNA landing region (Edward H Egelman et al. 1995). This means that the ring must be transiently opened in order for it to encircle the DNA. If it is possible for the ring to open during loading, this may indicate that it is possible for the subunits to spontaneously separate during the translocation and unwinding process. It would be interesting to better understand both the mechanism of loading and what this mechanism reveals about the states of the ring during its normal function.

One hint which could be pursued further is the enormous increase in helicase concentration required when the magnesium concentration or NTP concentration is varied. For example, at a magnesium concentration of 7mM, only 2nM of helicase monomer is required to get typical loading times of 80s, whereas with 50nM magnesium monomer concentrations of 200nM are required. This indicates that the magnesium concentration seems to affect the propensity of the ring to open. It would be interesting to better understand the connection between the unwinding behavior of the helicase under these conditions and the loading behavior.

This could be probed by conducting a systematic study of helicase concentration required to initiate unwinding given a certain single stranded loading region as a function of buffer conditions. However, it may be that the helicase can load but fails to translocate. This cannot be directly observed with the current experimental design since the motion of the helicase can only be observed indirectly through the fork position.

An alternative experimental design would be to attempt to observe the change in stiffness as a helicase binds to the single stranded region. Since the DNA binding arms in the channel will bind to the ssDNA as the helicase loads, this will slightly alter the stiffness of the DNA construct. By calculating the change in stiffness as a function of applied force one could estimate the number of helicase proteins bound to the DNA.

This is a challenging experiment as the expected signal from the helicase binding will be very small. At best all six binding arms lock onto subsequent base on the backbone, and bring them into the same plane. This would remove 5nt of ssDNA from the extension. At a force of 15pN this would result in a change in length of around 2.5nm. However, this is likely to be a gross overestimate since it is unlikely that either all the binding arms link to the DNA, and also that they are brought to the same plane, given the structural evidence (Singleton et al. 2000; Enemark & Joshua-Tor 2006).

Although this is a small signal there is also a major advantage when compared to the DNA unwinding experiment, since the measurement bandwidth can be dramatically decreased. If, for example, a helicase concentration is chosen so that there is one binding event every two minutes, the measurement bandwidth can be reduced to less than 0.01Hz. This means that a large amount of noise can be averaged out.

These very low bandwidths are possible because of the high stability of the dual trap configuration. In a single trap experiment the low frequency stage drift would make this impossible to perform.

Eukaryotic helicases

The ultimate goal of this work is to investigate and understand the physical mechanism of the more complex eukaryotic helicases. This helicase is called MCM2-7 and forms a hexamer, similar to T7 helicase, but with non-identical subunits (Y. Adachi et al. 1997). This helicase is responsible for separating the dsDNA of the genome during cell replication in eukaryotes.

As with many eukaryotic analogs to prokaryotic proteins, the MCM2-7 system is far more complicated in structure and function, and is significantly more challenging to assemble in vitro. It is only recently that labs have begun to isolate the proteins required and to understand the structure of the protein (Bochman & Schwacha 2009).

There is no crystal structure of eukaryotic MCM. However, there have been some low resolution electron microscopy studies suggesting it forms a hexamer both for the protein in yeast (Y. Adachi et al. 1997) and humans (Bochman & Schwacha 2007), (Norikazu Yabuta et al. 2003) and (M Sato et al. 2000). Biochemical experiments also show that the stoichiometry of the subunits in yeast indicates a heterohexamer (Y. Adachi et al. 1997).

Determining the mechanism of the MCM2-7 complex would be a very important discovery. There is likely to be a wide variety of functionality due to variations in the makeup of the ring and interactions with many accessory proteins (Bochman & Schwacha 2009). This will indubitably prompt an enormous amount of experimental studies once the protein becomes available for *in vitro* work. Applying the techniques we have used to study T7gp4 to MCM helicases would reveal many important parameters of these eukaryotic replicative helicases.

Summary

There are several follow up experiments which should be performed to verify the step-size measurement. Varying the GC sequence is important to check for any sequence based steps, instead of those purely due to the protein mechanism, though care must be taken not to form G quadruplex structures. An AT sequence may also be used, although this is complicated by the low unzipping force. There are also several other extensions to this work for the future. One is to examine the loading mechanism of T7gp4 helicase by identifying changes in stiffness as the protein binds. This may be possible despite the small signal due to the stability of the dual trap allowing extremely low bandwidth measurements. In the very long term, the same techniques applied to T7gp4 may be applied to the eukaryotic MCM2-7 replicative helicase.

BIBLIOGRAPHY

Adachi, K. et al., 2000. Stepping rotation of F1-ATPase visualized through angle-resolved single-fluorophore imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13), 7243-7247.

Adachi, Y., Usukura, J. & Yanagida, M., 1997. A globular complex formation by Nda1 and the other five members of the MCM protein family in fission yeast. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms*, 2(7), 467-479.

Ahnert, P. & Patel, S.S., 1997. Asymmetric interactions of hexameric bacteriophage T7 DNA helicase with the 5'- and 3'-tails of the forked DNA substrate. *The Journal of Biological Chemistry*, 272(51), 32267-32273.

Alberts, B., 2007. *Molecular Biology of the Cell* 5th ed., Other.

Ashkin, A., 1970. Acceleration and Trapping of Particles by Radiation Pressure. *Physical Review Letters*, 24(4), 156.

Ashkin, A. et al., 1986. Observation of a single-beam gradient force optical trap for dielectric particles. *Optics Letters*, 11(5), 288-290.

Benkovic, S.J., Valentine, A.M. & Salinas, F., 2001. Replisome-mediated DNA replication. *Annual Review of Biochemistry*, 70, 181-208.

Bernstein, J.A. & Richardson, C.C., 1988. A 7-kDa region of the bacteriophage T7 gene 4 protein is required for primase but not for helicase activity. *Proceedings of the National Academy of Sciences of the United States of America*, 85(2), 396-400.

Betterton, M.D. & Jülicher, F., 2005. Opening of nucleic-acid double strands by helicases: active versus passive opening. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 71(1 Pt 1), 011904.

Betterton, M.D. & Julicher, F., 2003. A Motor that Makes Its Own Track: Helicase Unwinding of DNA. *Physical Review Letters*, 91(25), 258103.

Bochman, M.L. & Schwacha, A., 2007. Differences in the single-stranded DNA binding activities of MCM2-7 and MCM467: MCM2 and MCM5 define a slow ATP-dependent step. *The Journal of Biological Chemistry*, 282(46), 33795-33804.

Bochman, M.L. & Schwacha, A., 2009. The Mcm Complex: Unwinding the Mechanism of a Replicative Helicase. *Microbiol. Mol. Biol. Rev.*, 73(4), 652-683.

Brewer, L. & Bianco, P., 2008. Laminar flow cells for single-molecule studies of DNA-protein interactions. *NATURE METHODS*, 5(6), 517-525.

Bryan, T.M. & Baumann, P., 2010. G-quadruplexes: from guanine gels to chemotherapeutics. *Methods in Molecular Biology (Clifton, N.J.)*, 608, 1-16.

Carter, A.R. et al., 2007. Stabilization of an optical microscope to 0.1 nm in three dimensions. *Applied Optics*, 46(3), 421-427.

Carter, B.C., Vershinin, M. & Gross, S.P., 2008. A Comparison of Step-Detection Methods: How Well Can You Do? *Biophysical Journal*, 94(1), 306–319.

Costa, A. & Onesti, S., 2008. The MCM complex: (just) a replicative helicase? *Biochemical Society Transactions*, 36(Pt 1), 136-40.

Crampton, D.J. et al., 2006. Oligomeric states of bacteriophage T7 gene 4 primase/helicase. *Journal of Molecular Biology*, 360(3), 667-677.

Donmez, I. & Patel, S.S., 2008. Coupling of DNA unwinding to nucleotide hydrolysis in a ring-shaped helicase. *The EMBO Journal*, 27(12), 1718-1726.

Egelman, E.H. et al., 1995. Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases. *Proceedings of the National Academy of Sciences of the United States of America*, 92(9), 3869-3873.

Egelman, E.H. et al., 1995. Bacteriophage T7 Helicase/Primase Proteins Form Rings Around Single-Stranded DNA that Suggest a General Structure for Hexameric Helicases. Available at: http://adsabs.harvard.edu/abs/1995PNAS...92.3869E [Accessed May 25, 2010].

Enemark, E.J. & Joshua-Tor, L., 2006. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature*, 442(7100), 270-275.

Fallman, E. & Axner, O., 2003. Influence of a Glass-Water Interface on the On-Axis Trapping of Micrometer-Sized Spherical Objects by Optical Tweezers. *Applied Optics*, 42(19), 3915-3926.

Gai, D. et al., 2004. Mechanisms of Conformational Change for a Replicative Hexameric Helicase of SV40 Large Tumor Antigen. *Cell*, 119(1), 47-60.

Gittes, F. & Schmidt, C.F., 1998. Interference model for back-focal-plane displacement detection in optical tweezers. *Optics Letters*, 23(1), 7-9.

Gorbalenya, A.E. & Koonin, E.V., 1993. Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology*, 3(3), 419-429.

Grossman, L. et al., 1988. Repair of DNA-containing pyrimidine dimers. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 2(11), 2696-2701.

Hecht, E., 2001. *Optics* 4th ed., Addison Wesley.

Howard-Flanders, P., 1968. DNA repair. *Annual Review of Biochemistry*, 37, 175-200.

Jeong, Y., Levin, M.K. & Patel, S.S., 2004. The DNA-unwinding mechanism of the ring helicase of bacteriophage T7. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19), 7264-7269.

Johnson, A. & O'Donnell, M., 2005. CELLULAR DNA REPLICASES: Components and Dynamics at the Replication Fork. *Annual Review of Biochemistry*, 74(1), 283-315.

Johnson, D.S. et al., 2007. Single-Molecule Studies Reveal Dynamics of DNA Unwinding by the Ring-Shaped T7 Helicase. *Cell*, 129(7), 1299-1309.

Kaplan, D.L. & O'Donnell, M., 2002. DnaB Drives DNA Branch Migration and Dislodges Proteins While Encircling Two DNA Strands. *Molecular Cell*, 10(3), 647-657.

Keniry, M.A., 2000. Quadruplex structures in nucleic acids. *Biopolymers*, 56(3), 123-146.

Kerssemakers, J.W.J. et al., 2006. Assembly dynamics of microtubules at molecular resolution. *Nature*, 442(7103), 709-712.

Kienberger, F. et al., 2005. Single molecule studies of antibody-antigen interaction strength versus intra-molecular antigen stability. *Journal of Molecular Biology*, 347(3), 597-606.

Kim, D., Narayan, M. & Patel, S.S., 2002. T7 DNA Helicase: A Molecular Motor that Processively and Unidirectionally Translocates Along Single-stranded DNA. *Journal of Molecular Biology*, 321(5), 807-819.

Koch, S. et al., 2002. Probing Protein-DNA Interactions by Unzipping a Single DNA Double Helix. *Biophysical Journal*, 83(2), 1098-1105.

Korpel, A., 1996. *Acousto-optics* 2nd ed., CRC Press.

Landry, M.P. et al., 2009. Characterization of photoactivated singlet oxygen damage in single-molecule optical trap experiments. *Biophysical Journal*, 97(8), 2128-2136.

Lee, J. et al., 1998. Coordinated leading and lagging strand DNA synthesis on a minicircular template. *Molecular Cell*, 1(7), 1001-1010.

Liao, J. et al., 2005. Mechanochemistry of T7 DNA Helicase. *Journal of Molecular Biology*, 350(3), 452-475.

Lohman, T.M. & Bjornson, K.P., 1996. Mechanisms of helicase-catalyzed DNA unwinding. *Annual Review of Biochemistry*, 65, 169-214.

Lynch, S. et al., 2009. Single Molecule Force Spectroscopy on G-Quadruplex DNA. *Chemistry - A European Journal*, 15(33), 8113-8116.

Mahamdeh, M. & Schaffer, E., 2009. Optical tweezers with millikelvin precision of temperature-controlled objectives and base-pair resolution. *Optics Express*, 17(19), 17190-17199.

Matson, S.W. & Richardson, C.C., 1983. DNA-dependent nucleoside 5'-triphosphatase activity of the gene 4 protein of bacteriophage T7. *Journal of Biological Chemistry*, 258(22), 14009-14016.

Meiners, J. & Quake, S.R., 1999. Direct Measurement of Hydrodynamic Cross Correlations between Two Particles in an External Potential. *Physical Review Letters*, 82(10), 2211.

Mendelman, L.V. & Richardson, C.C., 1991. Requirements for primer synthesis by bacteriophage T7 63-kDa gene 4 protein. Roles of template sequence and T7 56-kDa gene 4 protein. *Journal of Biological Chemistry*, 266(34), 23240-23250.

Miyata, T. et al., 2000. Two different oligomeric states of the RuvB branch migration motor protein as revealed by electron microscopy. *Journal of Structural Biology*, 131(2), 83-89.

Moffitt, J.R. et al., 2006. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proceedings of the National Academy of Sciences*, 103(24), 9006-9011.

Nelson, D.L. & Cox, M.M., 2004. *Lehninger Principles of Biochemistry, Fourth Edition* Fourth Edition., W. H. Freeman.

Nelson, P.C. et al., 2006. Tethered Particle Motion as a Diagnostic of DNA Tether Length. *The Journal of Physical Chemistry B*, 110(34), 17260-17267.

Neuman, K.C. et al., 2005. Statistical determination of the step size of molecular motors. *Journal of Physics: Condensed Matter*, 17(47), S3811-S3820.

Neuman, K.C. & Block, S.M., 2004. Optical trapping. *Review of Scientific Instruments*, 75(9), 2787.

Nugent-Glandorf, L. & Perkins, T.T., 2004. Measuring 0.1-nm motion in 1 ms in an optical microscope with differential back-focal-plane detection. *Optics Letters*, 29(22), 2611-2613.

Odijk, T., 1995. Stiff Chains and Filaments under Tension. *Macromolecules*, 28(20), 7016-7018.

Patel, S.S. & Hingorani, M.M., 1993. Oligomeric structure of bacteriophage T7 DNA primase/helicase proteins. *Journal of Biological Chemistry*, 268(14), 10668-10675.

Patel, S. & Picha, K., 2000. *Structure and function of hexameric helicases*, Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0033786801&partnerID=40 [Accessed April 4, 2009].

Patel, S.S. & Donmez, I., 2006. Mechanisms of Helicases. *Journal of Biological Chemistry*, 281:, 18265 -18268.

Pawley, J., 1995. *Handbook of Biological Confocal Microscopy* 2nd ed., Springer.

Picha, K.M. & Patel, S.S., 1998. Bacteriophage T7 DNA Helicase Binds dTTP, Forms Hexamers, and Binds DNA in the Absence of Mg2+. THE PRESENCE OF dTTP IS SUFFICIENT FOR HEXAMER FORMATION AND DNA BINDING. *J. Biol. Chem.*, 273(42), 27315-27319.

Pomerantz, R. & Odonnell, M., 2007. Replisome mechanics: insights into a twin DNA polymerase machine. *Trends in Microbiology*, 15(4), 156-164.

Puri, T. et al., 2007. Dodecameric Structure and ATPase Activity of the Human TIP48/TIP49 Complex. *Journal of Molecular Biology*, 366(1), 179-192.

Rogers, G.W., Komar, A.A. & Merrick, W.C., 2002. eIF4A: the godfather of the DEAD box helicases. *Progress in Nucleic Acid Research and Molecular Biology*, 72, 307-331.

Saikrishnan, K. et al., 2009. Mechanistic basis of 5'-3' translocation in SF1B helicases. *Cell*, 137(5), 849-859.

Sato, M. et al., 2000. Electron microscopic observation and single-stranded DNA binding activity of the Mcm4,6,7 complex. *Journal of Molecular Biology*, 300(3), 421-431.

Sawaya, M.R. et al., 1999. Crystal Structure of the Helicase Domain from the Replicative Helicase-Primase of Bacteriophage T7. *Cell*, 99(2), 167-177.

Scherzinger, E. & Seiffert, D., 1975. Studies on bacteriophage T7 DNA synthesis in vitro. I. Resolution of the T7 replication system into its components. *Molecular & General Genetics: MGG*, 141(3), 213-232.

Singleton, M.R., Dillingham, M.S. & Wigley, D.B., 2007. Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry*, 76, 23-50.

Singleton, M.R. et al., 2000. Crystal Structure of T7 Gene 4 Ring Helicase Indicates a Mechanism for Sequential Hydrolysis of Nucleotides. *Cell*, 101(6), 589-600.

Singleton, M.R. & Wigley, D.B., 2003. Multiple roles for ATP hydrolysis in nucleic acid modifying enzymes. *EMBO J*, 22(18), 4579-4583.

Smith, S.B., Cui, Y. & Bustamante, C., 1996. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science*, 271(5250), 795-799.

Stano, N.M. et al., 2005. DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase. *Nature*, 435(7040), 370-373.

Story, R.M., Weber, I.T. & Steitz, T.A., 1992. The structure of the E. coli recA protein monomer and polymer. *Nature*, 355(6358), 318-325.

Strunz, T. et al., 1999. Dynamic force spectroscopy of single DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 96(20), 11277-11282.

Thomsen, N.D. & Berger, J.M., 2009. Running in Reverse: The Structural Basis for Translocation Polarity in Hexameric Helicases. *Cell*, 139(3), 523-534.

Tolić-Norrelykke, S.F. et al., 2006. Calibration of optical tweezers with positional detection in the back focal plane. *Review of Scientific Instruments*, 77(10), 103101.

Toth, E.A. et al., 2003. The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Molecular Cell*, 12(5), 1113-1123.

Tyynismaa, H. et al., 2004. Twinkle helicase is essential for mtDNA maintenance and regulates mtDNA copy number. *Hum. Mol. Genet.*, 13(24), 3219-3227.

Wang, M.D. et al., 1997. Stretching DNA with optical tweezers. *Biophysical Journal*, 72(3), 1335–1346.

Yabuta, N. et al., 2003. Mammalian Mcm2/4/6/7 complex forms a toroidal structure. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms*, 8(5), 413-421.

Ye, J. et al., 2004. RecA-like motor ATPases--lessons from structures. *Biochimica Et Biophysica Acta*, 1659(1), 1-18.